

# Resolving the Geometric Locus Dilemma for Support Vector Learning Machines

Denise M. Reeves

## Abstract

Capacity control, the bias/variance dilemma, and learning unknown functions from data, are all concerned with identifying effective and consistent fits of unknown geometric loci to random data points. A geometric locus is a curve or surface formed by points, all of which possess some uniform property. A geometric locus of an algebraic equation is the set of points whose coordinates are solutions of the equation. Any given curve or surface must pass through each point on a specified locus. This paper argues that it is impossible to fit random data points to algebraic equations of partially configured geometric loci that reference arbitrary Cartesian coordinate systems. It also argues that the fundamental curve of a linear decision boundary is actually a principal eigenaxis. It is shown that learning principal eigenaxes of linear decision boundaries involves finding a point of statistical equilibrium for which eigenenergies of principal eigenaxis components are symmetrically balanced with each other. It is demonstrated that learning linear decision boundaries involves strong duality relationships between a statistical eigenlocus of principal eigenaxis components and its algebraic forms, in primal and dual, correlated Hilbert spaces. Locus equations are introduced and developed that describe principal eigen-coordinate systems for lines, planes, and hyperplanes. These equations are used to introduce and develop primal and dual statistical eigenlocus equations of principal eigenaxes of linear decision boundaries. Important generalizations for linear decision boundaries are shown to be encoded within a dual statistical eigenlocus of principal eigenaxis components. Principal eigenaxes of linear decision boundaries are shown to encode Bayes' likelihood ratio for common covariance data and a robust likelihood ratio for all other data.

**Keywords:** geometric locus dilemma, strong dual principal eigenlocus method, critical minimum eigenenergy, statistical eigenlocus, principal eigenaxis, normal eigenaxis, strong dual normal eigenlocus identity, strong dual normal eigenlocus likelihood ratio, Bayes' likelihood ratio, statistical equilibrium, probabilistic multiclass linear classifier, probabilistic binary linear classifier, bias/variance dilemma, capacity control, support vector learning machines.

# Contents

<b>1</b>	<b>Introduction and Motivation</b>	<b>6</b>
1.1	Paradigm Shift in Machine Learning: Targeting the Right Curves	7
1.2	The Paradigm Shift . . . . .	8
1.3	Learning the Locus of a Principal Eigenaxis . . . . .	8
1.4	Organization of the Paper . . . . .	9
<b>2</b>	<b>Fitting Learning Machine Architectures to Unknown Functions of Data</b>	<b>11</b>
2.1	Regulation of Learning Machine Capacities . . . . .	12
2.2	SVM Capacity Control . . . . .	13
2.3	The Bias & Variance Dilemma . . . . .	14
2.3.1	Essence of the Bias/Variance Dilemma . . . . .	15
2.4	Prewiring of Important Generalizations . . . . .	15
2.5	The System Representation Problem . . . . .	16
2.6	Suitable Representations for Learning Machine Architectures . .	17
2.7	Tractable Statistical Models for SVM Architectures . . . . .	17
2.8	A Tractable Dual Locus of Eigen-scaled Data Points . . . . .	18
2.8.1	Partially Specified Principal Eigenaxes . . . . .	18
2.8.2	Properly Specified Normal Eigenaxes . . . . .	18
2.9	The Graph or Locus of an Equation . . . . .	19
2.10	Methods for Solving Locus Problems . . . . .	19
2.11	Changing the Loci of Reference Axes . . . . .	20
2.12	Geometric Locus of a Straight Line . . . . .	20
<b>3</b>	<b>The Geometric Locus Dilemma</b>	<b>20</b>
3.1	An Impossible Estimation Task . . . . .	21
3.2	SVMs and the Geometric Locus Dilemma . . . . .	21
3.2.1	Linear Interpolation Using Random Slack Variables . . .	22
3.3	Ill-defined SVM Architectures . . . . .	23
3.4	How Does Linear Kernel SVM Learn from Data? . . . . .	24
3.4.1	Classification Example One . . . . .	24
3.4.2	Classification Example Two . . . . .	25
<b>4</b>	<b>An Elegant Principal Eigen-coordinate System for Linear Loci</b>	<b>27</b>
4.1	Locus Equations of a Normal Eigenaxis . . . . .	27
4.2	The Normal Eigenaxis of a Linear Locus . . . . .	29
4.3	Fundamental Equation of a Linear Locus . . . . .	30
4.4	Major Intrinsic Axes of Second-order Loci . . . . .	30
4.5	The Principal Eigenaxis of a Linear Locus . . . . .	31
4.6	Coordinate Form Locus Equation of a Unit Normal Eigenaxis . .	32
4.7	Uniform Property Exhibited by Points on a Linear Locus . . . .	33
4.8	Properties of Normal Eigenaxes . . . . .	34
4.8.1	Characteristic Eigenenergy of a Normal Eigenaxis . . . .	34

4.9	Correlated Uniform Properties Exhibited by Points on a Linear Locus . . . . .	34
<b>5</b>	<b>Design of Learning Machine Architectures in Hilbert Spaces</b>	<b>36</b>
5.1	Geometric Properties of a Vector . . . . .	36
5.2	Geometric Locus of a Vector . . . . .	38
5.3	Uniform Property Exhibited by Vector Components . . . . .	38
5.4	Inner Product Statistics . . . . .	39
5.5	Second-order Distance Statistics Between Loci of Vectors . . . .	39
5.6	Signed Magnitudes of Vector Projections . . . . .	41
5.7	Designing Functional Glue for Learning Machine Architectures .	42
5.8	Hardwiring the Eigenlocus of a Normal Eigenaxis into Linear Kernel SVM Architectures . . . . .	44
5.9	A Strong Dual Normal Eigenlocus of Normal Eigenaxis Components	45
5.10	High Level Overview of a Strong Dual Normal Eigenlocus . . . .	45
<b>6</b>	<b>The Primal and the Wolfe Dual Normal Eigenlocus Equations of a Probabilistic Binary Linear Classification System</b>	<b>46</b>
6.1	Eigenlocus Equation of a Primal Normal Eigenlocus . . . . .	47
6.2	The Critical Minimum Eigenenergy Constraint on $\tau$ . . . . .	48
6.3	The Strong Dual Normal Eigenlocus Equilibrium Point . . . . .	49
6.4	Why the Wolfe Dual Normal Eigenlocus Matters . . . . .	49
6.5	Fundamental Unknowns for Strong Dual Normal Eigenlocus Estimates . . . . .	50
6.6	The Wolfe Dual Normal Eigenlocus of a Separating Hyperplane .	51
6.7	Symmetrical Linear Partitioning Systems in $\mathbb{R}^N$ and $\mathbb{R}^d$ . . . .	52
6.8	Strong Duality Relationships Between a Constrained Primal and a Wolfe Dual Normal Eigenlocus . . . . .	53
6.9	Uniform Geometric and Statistical Properties Jointly Exhibited by Correlated Normal Eigenaxis Components on $\tau$ and $\psi$ . . . .	54
6.10	Fundamental Relationships Between Joint Statistical Estimates of $\tau$ and $\psi$ . . . . .	54
<b>7</b>	<b>The Lagrangian of the Primal Normal Eigenlocus</b>	<b>55</b>
7.1	A Statistical Decision System for Probabilistic Binary Linear Classification . . . . .	56
7.1.1	Eigenlocus Equation of the Linear Decision Boundary . .	57
7.1.2	Eigenlocus Equation of the $D_{+1}(\mathbf{x})$ Decision Border . . .	57
7.1.3	Eigenlocus Equation of the $D_{-1}(\mathbf{x})$ Decision Border . . .	57
7.2	Axis of Symmetry for Bilateral Linear Partitions . . . . .	59
<b>8</b>	<b>Statistical Representation of <math>\tau</math> Within the Wolfe Dual Eigenspace</b>	<b>59</b>
8.1	The Constrained Primal Normal Eigenlocus . . . . .	59
8.1.1	Extreme Data Points of Overlapping Data Distributions .	61
8.1.2	Extreme Data Points of Non-overlapping Data Distributions	61
8.2	The Pair of Strong Dual Normal Eigenlocus Components . . . .	62

<b>9</b>	<b>Width Regulation of Linear Decision Regions</b>	<b>63</b>
9.1	Bipartite Symmetric Partitions of Large Covariance Regions . . .	65
9.2	Strong Dual Normal Eigenlocus Transforms for Non-overlapping Data Distributions . . . . .	65
9.3	Beyond Classical Interpolation Methods . . . . .	66
9.4	Strong Dual Normal Eigenlocus Transforms for Overlapping Data Distributions . . . . .	69
9.5	Regularized and Customized Geometric Architectures . . . . .	72
9.6	Balancing the Total Allowed Eigenenergies of $\tau_1$ and $\tau_2$ . . . . .	74
9.7	KKT Complementary Conditions . . . . .	74
9.8	Statistical Functionality of the $\tau_0$ Term . . . . .	75
9.9	The Normal Eigenlocus Test Statistic . . . . .	75
9.9.1	Statistical Decision Locus . . . . .	76
9.10	Strong Dual Normal Eigenlocus Transforms for Homogeneous Distributions . . . . .	78
9.11	Equilibrium States of Strong Dual Decision Systems . . . . .	79
9.12	Critical Minimum Eigenenergy Constraints on $\tau_1$ , $\tau_2$ , and $\tau$ . . .	80
9.13	Equilibrium Constraints on Wolfe Dual Normal Eigenaxis Components . . . . .	80
<b>10</b>	<b>Eigenlocus Equation of a Wolfe Dual Normal Eigenlocus</b>	<b>81</b>
10.1	The Wolfe Dual Normal Eigenlocus . . . . .	83
10.1.1	Second-degree Homogeneous Polynomial Surfaces . . . . .	83
10.1.2	Geometric Essence of $\psi$ . . . . .	83
10.2	Fundamental Properties of $\psi$ . . . . .	84
10.2.1	The Critical Minimum Eigenenergy Constraint on $\psi$ . . .	84
10.3	Wolfe Dual Statistical Systems of Partitioning Hyperplanes . . .	85
<b>11</b>	<b>Weak Dual Normal Eigenlocus Transforms</b>	<b>87</b>
11.1	Eigenspectrums of Gram Matrices . . . . .	87
11.2	Incomplete Eigenspectrums of Low Rank Gram Matrices . . . . .	87
11.3	Principal Statistical States and Characteristic Eigenstates of Strong Dual Decision Systems . . . . .	88
11.4	Generating Sufficient Eigenspectrums for Low Rank Gram Matrices	90
11.5	Eigenspectrum Shaping of Quadratic Surfaces . . . . .	91
11.6	Descriptive Statistics Encoded Within $\psi$ . . . . .	94
<b>12</b>	<b>Point and Coordinate Relationships Between Constrained Primal and Wolfe Dual Normal Eigenaxis Components</b>	<b>95</b>
12.1	Joint Statistical Underpinnings of $\psi$ and $\tau$ . . . . .	96
12.2	Distributions of First Degree Vector Coordinates . . . . .	97
12.2.1	Signed Magnitudes of Vector Projections . . . . .	98
12.3	Omnidirectional and Unidirectional Covariance Statistics . . . .	100
12.3.1	Omnidirectional Covariance Statistics . . . . .	100
12.3.2	Pointwise Covariance Statistics . . . . .	101

12.4	Discovery of Extreme Data Points with Pointwise Covariance Statistics . . . . .	102
12.5	Eigen-scaled Pointwise Covariance Statistics . . . . .	104
12.6	Descriptive Statistics for Separating Hyperplanes . . . . .	105
12.7	Symmetrical Relationships Between the Total Allowed Eigenenergies of a Strong Dual Normal Eigenlocus . . . . .	105
<b>13</b>	<b>Underneath the Hood of a Wolfe Dual Normal Eigenlocus</b>	<b>107</b>
13.1	Non-Orthogonal Eigenaxes of $\psi$ . . . . .	109
<b>14</b>	<b>Eigenloci of the <math>\psi_{1i*} \vec{e}_{1i*}</math> Wolfe Dual Normal Eigenaxis Components</b>	<b>111</b>
14.1	Uniform Geometric and Statistical Properties Jointly Exhibited by Normal Eigenaxis Components on $\psi$ and $\tau_1$ . . . . .	113
14.2	Directional Symmetries Exhibited by Normal Eigenaxis Components on $\psi$ and $\tau_1$ . . . . .	115
<b>15</b>	<b>Eigenloci of the <math>\psi_{2i*} \vec{e}_{2i*}</math> Wolfe Dual Normal Eigenaxis Components</b>	<b>118</b>
15.1	Uniform Geometric and Statistical Properties Jointly Exhibited by Normal Eigenaxis Components on $\psi$ and $\tau_2$ . . . . .	120
15.2	Directional Symmetries Exhibited by Normal Eigenaxis Components on $\psi$ and $\tau_2$ . . . . .	121
<b>16</b>	<b>Properties Exhibited by the Total Allowed Eigenenergy of a Wolfe Dual Normal Eigenlocus <math>\psi</math></b>	<b>125</b>
16.1	Finding the Right Component Lengths . . . . .	126
16.2	Critical Length Constraints . . . . .	127
<b>17</b>	<b>An Elegant Statistical Balancing Feat</b>	<b>128</b>
17.1	The Law of Cosines for Strong Dual Normal Eigenlocus Transforms	130
17.2	Examining the Total Allowed Eigenenergies of a Strong Dual Normal Eigenlocus . . . . .	131
17.2.1	The Total Allowed Eigenenergy of $\tau_1$ . . . . .	131
17.2.2	The Total Allowed Eigenenergy of $\tau_2$ . . . . .	133
17.2.3	The Total Allowed Eigenenergy of $\tau$ . . . . .	134
17.3	Balancing the Total Allowed Eigenenergies of $\tau_1$ and $\tau_2$ . . . . .	135
17.4	The State of Statistical Equilibrium of a Strong Dual Normal Eigenlocus . . . . .	137
17.5	The Statistical Machinery Behind the Balancing Feat . . . . .	137
17.6	Statistical Machinery of a Statistical Fulcrum and a Statistical Lever . . . . .	138
17.7	Characteristics of the State of Statistical Equilibrium . . . . .	139
17.8	The Strong Dual Normal Eigenlocus Identity . . . . .	141

<b>18 Probabilistic Properties Exhibited by the Statistical Equilibrium Point of a Strong Dual Normal Eigenlocus</b>	<b>147</b>
18.1 A Probabilistic Explanation for the Total Allowed Eigenenergies of $\tau$ . . . . .	148
18.2 Likelihood Ratios Encoded Within $\tau$ . . . . .	149
18.2.1 Equivalence of Class Probabilities . . . . .	151
18.3 Probabilistic Expressions of Decision Region Widths . . . . .	152
18.4 The Normal Eigenlocus Likelihood Ratio . . . . .	153
18.5 Comparison of the Normal Eigenlocus Decision Rule with Bayes' Decision Rule . . . . .	154
18.6 Previous Work on Linear Kernel SVMs . . . . .	156
<b>19 Design of Probabilistic Multiclass Linear Pattern Recognition Systems</b>	<b>158</b>
19.1 Design of Customized Probabilistic Statistical Decision Engines .	159
19.2 Practical Dual-Use of Strong Dual Normal Eigenlocus Discriminant Functions . . . . .	160
19.3 Using Normal Eigenlocus Test Statistics to Design Effective Feature Extractors . . . . .	161
19.3.1 A Robust Statistical Multimeter . . . . .	161
19.4 Summary of Practical Dual-Use of Linear Kernel SVMs . . . . .	162
<b>20 Synopsis of Geometric Underpinnings and Statistical Machinery of Linear Kernel SVMs</b>	<b>162</b>
<b>21 Conclusions</b>	<b>165</b>

## 1 Introduction and Motivation

The design and development of learning machine architectures has primarily been based on curve and surface fitting methods of interpolation or regression, alongside statistical methods of reducing data to minimum numbers of relevant parameters. For example, multilayer artificial neural networks (ANNs) estimate nonlinear regressions with optimally pruned architectures. Good generalization performance for ANNs is considered an effect of a good nonlinear interpolation of the training data Geman et al. [1992], Haykin [2009]. Alternatively, support vector machines (SVMs) fit linear curves or surfaces to minimum numbers of training points. Good generalization performance for SVMs is largely attributed to maximally separated linear decision borders Boser et al. [1992], Cortes and Vapnik [1995], Bennett and Campbell [2000], Cristianini and Shawe-Taylor [2000], Scholkopf and Smola [2002].

Most significant machine learning problems are considered extrapolation problems for unknown functions, e.g., nontrivial black box estimates. Black boxes are defined in terms of inputs, subsequent outputs, and the mathematical functions that relate them. Because training points will never cover a space of

possible inputs, practical learning machines must extrapolate in manners that provide effective generalization performance Geman et al. [1992], Gershenfeld [1999], Haykin [2009].

Identifying effective methods that provide consistent fits of unknown functions to random data points remains a difficult and open problem. This paper assumes that classical numerical methods for curve fitting and function approximation provide little insight into effective designs for learning machine architectures.

## 1.1 Paradigm Shift in Machine Learning: Targeting the Right Curves

This paper intends to introduce a paradigm shift in machine learning of curves or surfaces from data. The paradigm concerns the *strong dual principal eigenlocus method*. Principal eigenlocus methods involve learning principal eigenaxes of second-order decision boundaries that take the form of  $d$ -dimensional circles, ellipses, hyperbolae, parabolas, or lines. This paper will argue that learning second-order decision boundaries from training data requires learning the geometric locus of a principal eigenaxis. To wit, the primary curve of interest for second-order decision boundaries is the locus of a principal eigenaxis.

The fundamental idea behind the paradigm shift is the simple general concept of a geometric locus of points. A geometric locus is a curve or surface formed by a set of points, all of which possess some uniform property. Any given geometric locus is described by an algebraic equation, where the geometric locus of an algebraic equation is the locus (location) of all those points whose coordinates are solutions of the equation. Any point whose coordinate locations do not satisfy the algebraic equation of a specified locus is not on the given curve or surface. Likewise, any given curve or surface must pass through each point on a specified locus. Classic examples of a geometric locus include circles, ellipses, hyperbolae, parabolas, lines, and points. Material on geometric locus methods can be found in Nichols [1893], Tanner and Allen [1898], Whitehead [1911], and Eisenhart [1939].

For any given set of points which possess a uniform property, there exists a geometric locus, formed by those points which exhibit the uniform property, that is described by an algebraic equation, the solutions of which determine the locations of the points on the given locus.

This paper will examine effective designs for learning machine architectures that involve identifying and exploiting algebraic systems of geometric loci which provide *regularized geometric architectures of statistical decision systems*. Accordingly, this paper will consider effective designs for learning machine architectures that involve *targeting the right curves and learning the right set of geometric loci*.

## 1.2 The Paradigm Shift

Generally speaking, learning an unknown function from data involves finding the function that generated a set of points. *This paper will show that the locus of a principal eigenaxis is the place of origin for all forms of separating lines, planes, and hyperplanes.* This paper will develop the algebraic equations of the geometric locus of the principal eigenaxis of a linear locus of points. These equations will be used to identify the uniform geometric properties which are exhibited by all of the points on a linear locus, which include the principal eigenaxis of the locus. All of these results will be used to motivate and develop an algebraic system of locus equations that describes principal eigenaxes of linear decision boundaries. Moreover, the algebraic system of locus equations determines principal eigenaxes of separating lines, planes, or hyperplanes for all forms of data distributions, including overlapping and homogeneous distributions. Thereby, it will be shown that learning linear decision boundaries from training data essentially involves learning the locus of a principal eigenaxis.

## 1.3 Learning the Locus of a Principal Eigenaxis

The locus of a principal eigenaxis is an inherent part of any linear curve or surface. Moreover, the geometric locus of a linear curve or surface is encoded within the locus of its principal eigenaxis. It will be shown that the eigen-coordinate locations of a principal eigenaxis determine the uniform properties exhibited by the points on a linear curve or surface. It will be demonstrated that a principal eigenaxis satisfies a linear locus in terms of its eigenenergy. It will also be shown that principal eigenaxes of linear loci provide exclusive and distinctive reference axes.

A principal eigenaxis is considered to be a *characteristic locus* for a linear locus of points. Accordingly, the term *eigenlocus* will be used to refer to the principal eigenaxis of a linear locus. The term *statistical eigenlocus* will be used to refer to a statistical estimate of an eigenlocus of a linear decision boundary. The term *strong dual principal eigenlocus* will be used to refer to joint, statistical eigenlocus estimates in dual, correlated Hilbert spaces.

This paper will motivate and develop a strong dual principal eigenlocus of principal eigenaxis components which encodes the eigen-coordinate locations of an unknown principal eigenaxis of a linear decision boundary. The paper will examine how important generalizations for linear decision boundaries are encoded within statistical representations of principal eigenaxis locations, whereby principal eigenaxis locations describe statistical properties of training data. The paper will demonstrate how statistical representations of principal eigenaxes take the form of a strong dual principal eigenlocus of principal eigenaxis components, where each component encodes an eigen-transformed principal location of large covariance.

This paper will examine how encoding relevant statistical aspects of principal eigenaxis locations within learning machine architectures requires an algebraic system of primal and dual statistical eigenlocus equations that properly specify



the loci of principal eigenaxis components for a given set of training data. The paper will consider how this algebraic system of eigenlocus equations involves strong duality relationships between a statistical eigenlocus of principal eigenaxis components and its algebraic forms, in primal and dual, correlated Hilbert spaces. It will be shown that learning a strong dual principal eigenlocus of principal eigenaxis components involves finding a point of statistical equilibrium for which the eigenenergies of eigenlocus components are symmetrically balanced with each other in relation to a centrally located statistical fulcrum. It will be demonstrated that a strong dual principal eigenlocus satisfies a linear decision boundary in terms of a critical minimum, i.e., a total allowed, eigenenergy. It will also be demonstrated that a strong dual principal eigenlocus encodes Bayes' likelihood ratio for similar covariance data distributions and a robust likelihood ratio for all other data distributions.

The findings presented in this paper will define substantial geometric architectures and statistical models for linear kernel SVMs. Findings for polynomial kernel SVMs will be presented in another paper. New and surprising information, discovered from the creation and analysis of statistical eigen-coordinate systems for linear and nonlinear second-order decision boundaries, is expected to provide significant insights into the identification and exploitation of effective kernel widths for Gaussian kernel SVMs.

This paper will show that linear kernel SVMs are a powerful and robust class of statistical learning machines which are useful for the design, development, and implementation of probabilistic, multiclass linear pattern recognition systems.

The work presented in this paper is motivated by the seminal works of Thomas Cover (1965) Cover [1965], Geman, Bienenstock, and Doursat (1992) Geman et al. [1992], Boser, Guyon, and Vapnik (1992) Boser et al. [1992], and Cortes and Vapnik (1995) Cortes and Vapnik [1995].

## 1.4 Organization of the Paper

The paper is organized as follows. The issues of fitting learning machine architectures to unknown functions of training data are laid out in Section 2. In addition, Section 2 considers the system representation problem for learning machine architectures, methods for solving locus problems, and a high level overview of a statistical model for linear kernel SVMs. Section 3 formulates the geometric locus dilemma for statistical learning machines and SVMs. Section 3 also motivates resolving the geometric locus dilemma for linear and polynomial kernel SVMs. Section 4 develops a principal eigen-coordinate system that describes all forms of linear loci. Principal eigenaxes of linear loci are given the name of normal eigenaxes. Section 5 motivates the design of learning machine architectures in Hilbert spaces, develops the notion of functional glue for learning machine architectures, and motivates hardwiring the eigenlocus of a principal, i.e., normal, eigenaxis into linear kernel SVM architectures. Section 6 defines the primal and the Wolfe dual normal eigenlocus equations of a probabilistic, linear binary classification system. Section 7 defines the Lagrangian of the primal normal eigenlocus. Section 8 begins the process of defining the

primal normal eigenlocus within the Wolfe dual eigenspace. Section 9 examines width regulation of large covariance linear decision regions, develops a normal eigenlocus test statistic for classifying unknown data, considers strong dual normal eigenlocus transforms for homogeneous data distributions, and examines equilibrium conditions for strong dual normal eigenlocus transforms. Section 10 examines the strong dual normal eigenlocus problem within the context of the eigenlocus equation of a Wolfe dual normal eigenlocus. Section 11 considers how geometric and statistical properties of strong dual normal eigenlocus transforms are sensitive to eigenspectrums of Gram matrices; Section 11 also considers how eigenspectrums of Gram matrices determine shapes of quadratic surfaces. Section 12 motivates the examination of point and coordinate relationships between the constrained primal and the Wolfe dual normal eigenaxis components. Pointwise covariance statistics are defined for individual training points, and are used to find extreme data points which possess large pointwise covariances. Section 12 also considers the total allowed eigenenergies of a strong dual normal eigenlocus. Section 13 develops a general expression for a principal eigen-decomposition that is used to examine point and coordinate relationships between the constrained primal and the Wolfe dual normal eigenaxis components. Sections 14 and 15 develop algebraic expressions for the eigenloci (the geometric locations) of the Wolfe dual normal eigenaxis components. These expressions are used to define uniform geometric and statistical properties which are jointly exhibited by Wolfe dual and constrained primal normal eigenaxis components. Section 16 outlines the fundamental issue that must be resolved to ensure that eigenenergies of normal eigenlocus components are symmetrically balanced with each other. Section 17 examines the algebraic, geometric, and statistical nature of the remarkable statistical balancing feat that is routinely accomplished by strong dual normal eigenlocus transforms. Algebraic and statistical expressions are developed for the total allowed eigenenergies of a strong dual normal eigenlocus. These expressions are used to define the statistical machinery behind the statistical balancing feat, and are used to derive the strong dual normal eigenlocus identity. Section 18 defines probabilistic properties which are exhibited by strong dual normal eigenlocus discriminant functions. An expression is obtained for the normal eigenlocus decision rule which encodes likelihood ratios, and is used to show that strong dual normal eigenlocus discriminant functions encode Bayes' likelihood ratio for common covariance data, and a robust likelihood ratio for all other data distributions. Section 19 outlines dual-use of strong dual normal eigenlocus discriminant functions which include probabilistic, multiclass linear pattern recognition systems, a statistical multimeter for measuring class separability and Bayes' error rate, and a robust indicator of homogeneous data distributions. Section 20 summarizes the geometric underpinnings and statistical machinery of linear kernel SVMs. Section 21 summarizes the major findings and conclusions of the paper.

## 2 Fitting Learning Machine Architectures to Unknown Functions of Data

Fitting unknown functions to collections of random or arbitrary data points has long been described as fraught with difficulties. Functions tend to be overly sensitive to point coordinate locations and small perturbations of the data. Minor changes in data point locations can produce largely different functions that vary widely in performance. Function values and performance are also affected by data quantities Synge [1957], Daniel and Wood [1979], Lancaster and Salkauskas [1986], Wahba [1987], Linz and Wang [2003].

Most function approximation methods hinge on the guarantee that a straight line can be passed through two points, a parabola through three, a cubic through four, and so on Davis [1963]. A large class of function approximation problems involve fitting a set of curves or surfaces to some representative data set Lancaster and Salkauskas [1986], Rao [2002], Linz and Wang [2003], Rahman [2004]. For example, the Fourier series fits sinusoids or complex exponentials to periodic signals Lathi [1998], whereas wavelets are used to decompose signals into low frequency and high frequency components Mertins [1999]. Generally speaking, curve fitting involves selecting a function that generated a set of points.

Classical approximation methods were initially developed to replace more complicated functions with simpler ones, such as polynomial functions and piecewise polynomials. The primary concerns for classical approximation methods are approximation errors and speed of convergence Davis [1963], Linz [1979], Keener [2000], Linz and Wang [2003].

Fitting learning machine architectures to unknown functions of training data involves numerous and difficult problems. Learning machine architectures are extremely sensitive to algebraic and topological structures that include functionals, reproducing kernels, parameter sets, constraint sets, regularization parameters, and eigenspectrums of data matrices Geman et al. [1992], Byun and Lee [2002], Haykin [2009], Reeves [2009], Reeves and Jacyna [2011]. For example, SVM architectures based on polynomial and Gaussian kernel matrices vary widely in terms of generalization performance Byun and Lee [2002], Eitrich and Lang [2006]. It has also been shown that regularization parameters of linear kernel SVMs determine SVM architectures that exhibit extreme variations of generalization performance Reeves and Jacyna [2011]. The generalization performance of multilayer ANNs also varies substantially with data samples and functional (hidden node) configurations Haykin [2009]. In addition to these difficulties, parameter and eigenspectrum estimates may be ill-defined Kay [1993], Moon and Stirling [2000], Reeves [2009], Reeves and Jacyna [2011]. Identifying the correct form of an equation for a statistical model is also a large concern Daniel and Wood [1979], Breiman [1991], Geman et al. [1992].

In general, learning algorithms that estimate decision boundaries for classification problems introduce four sources of error into the final classification system: (1) Bayes' error, (2) model error or bias, (3) estimation error or variance, and (4) computational error VanTrees [1968], Tikhovov and Arsenin [1977],

Goldberg [1979], Linz [1979], Wahba [1987], Fukunaga [1990], Geman et al. [1992], Mitchell [1997], Engl et al. [2000], Duda et al. [2001], Zhdanov [2002], Linz and Wang [2003], Haykin [2009]. Bayes' error defines an optimal error rate for any decision making task VanTrees [1968], Fukunaga [1990], Duda et al. [2001].

## 2.1 Regulation of Learning Machine Capacities

Learning machine architectures with  $N$  free parameters have a learning capacity to fit  $N$  data points, so that curves or surfaces can be made to pass through every data point. However, fitting all of the training data is generally considered bad statistical practice Wahba [1987], Breiman [1991], Geman et al. [1992], Barron et al. [1998], Cherkassky and Mulier [1998], Gershenfeld [1999], Duda et al. [2001], Hastie et al. [2001], Haykin [2009]. Random fluctuations (noise) in signals or images obscures information contained in training data. Learning machine architectures that correctly interpolate collections of noisy training points, fit the idiosyncrasies of the noise and are not expected to exhibit good generalization performance. Likewise, highly flexible architectures with indefinite parameter sets are said to overfit the training data Wahba [1987], Geman et al. [1992], Cherkassky and Mulier [1998], Gershenfeld [1999], Duda et al. [2001], Hastie et al. [2001], Scholkopf and Smola [2002], Haykin [2009].

Yet again, learning machine architectures that interpolate insufficient numbers of data points exhibit under-fitting difficulties. Architectures with too few parameters ignore both the noise and the meaningful behavior of the data. Parameterized architectures that cannot adequately describe a set of data samples are said to underfit the training data Wahba [1987], Cherkassky and Mulier [1998], Gershenfeld [1999], Scholkopf and Smola [2002], Haykin [2009]. Figure 1 depicts a cartoon of an underfitting, overfitting, and balanced fitting of a set of data points.

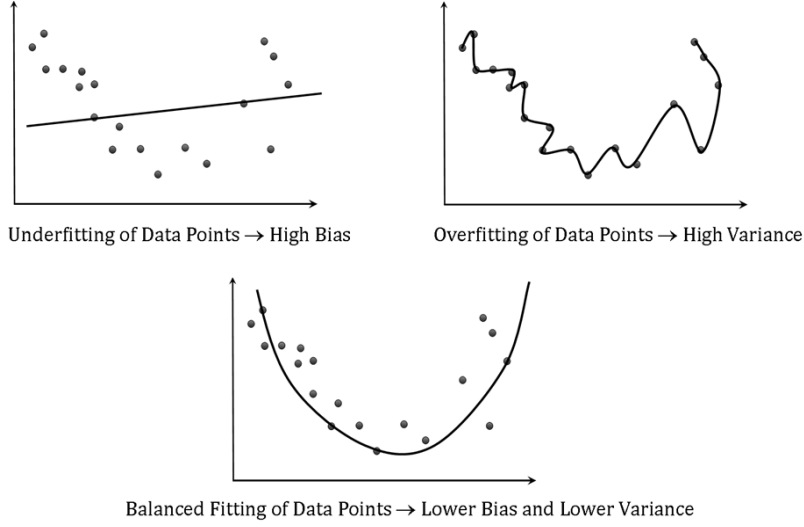


Figure 1: Illustration of the difficulties associated with fitting an unknown function to a collection of random data points.

Discriminant functions that possess too much learning capacity are more likely to overfit the training data Boser et al. [1992], Scholkopf and Smola [2002]. SVMs mitigate overfitting difficulties by a process termed capacity control which is outlined next.

## 2.2 SVM Capacity Control

SVMs estimate linear decision boundaries by solving a quadratic programming problem. The capacity or complexity of SVM decision boundaries is regulated by means of a geometric margin of separation between two given sets of data. The SVM method minimizes the capacity of a separating hyperplane by maximizing the distance between a pair of margin hyperplanes or linear borders. Large distances between margin hyperplanes (1) allow for considerably fewer hyperplane orientations, and (2) enforce a limited capacity to separate training data. Thus, maximizing the distance between margin hyperplanes regulates the complexity of separating hyperplane estimates Boser et al. [1992], Cortes and Vapnik [1995], Burges [1998], Bennett and Campbell [2000], Cristianini and Shawe-Taylor [2000], Scholkopf and Smola [2002]. Figure 2 illustrates a cartoon of three geometric margins for a collection of training data, where each linear border interpolates a data point from one of the pattern classes. The SVM method chooses the green linear borders which exhibit the largest geometric margin of separation.

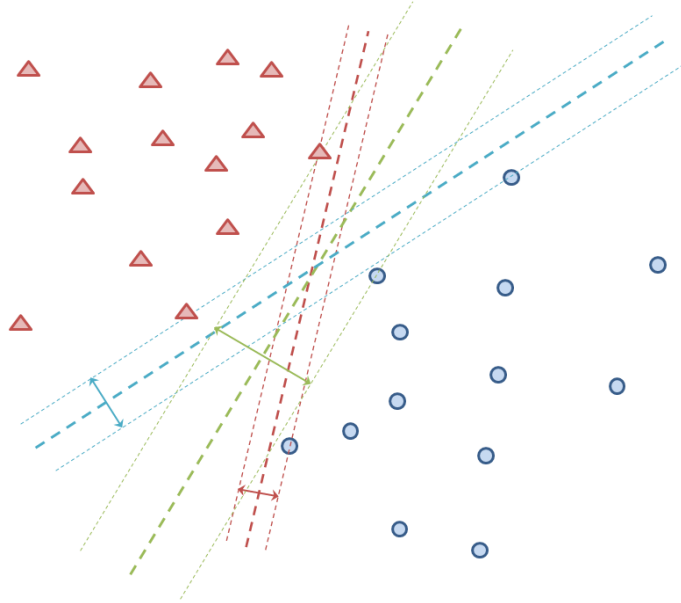


Figure 2: Illustration of three geometric margins of separation, where data points that belong to the same category are depicted by red triangles or blue circles. The green linear borders exhibit the largest geometric margin, whereas the red linear borders exhibit the smallest geometric margin.

All of the above difficulties imply that learning unknown functions from training data involves trade-offs between underfittings and overfittings of data points. The bias-variance dilemma describes statistical facets of these trade-offs Geman et al. [1992], Gershenfeld [1999], Duda et al. [2001], Hastie et al. [2001], Haykin [2009].

### 2.3 The Bias & Variance Dilemma

All learning machine architectures are composed of training data. Moreover, the estimation error between a learning machine and its target function depends on the training data in a twofold manner. Geman, Bienenstock, and Doursat examined these dual sources of estimation error in their seminal article titled *Neural Networks and the Bias/Variance Dilemma* Geman et al. [1992]. The crux of the dilemma is that estimation error is composed of two distinct components termed a bias and a variance. Large numbers of parameter estimates raise the variance, whereas incorrect statistical models increase the bias Geman et al. [1992], Gershenfeld [1999], Duda et al. [2001], Hastie et al. [2001], Haykin [2009].

Model-free approaches to learning, which are also known as nonparametric inference methods, assume no particular types of geometric structures, use large numbers of parameters, and exhibit high estimation variance. For example, learning machine architectures of multilayer ANNs, which are formed by

arbitrary sets and types of geometric surfaces, vary substantially with data samples and hidden node configurations. Such variance can only be reduced with sufficient amounts of data samples Geman et al. [1992], Haykin [2009]. On the other side of the dilemma, parametric or statistical models that use improper representations exhibit large modeling biases. For instance, learning Bayes' decision boundaries from training data drawn from Gaussian distributions involves the estimation of conic sections or quadratic surfaces. Discriminant functions that are based on incorrect statistical models do not achieve the Bayes' error rate VanTrees [1968], Fukunaga [1990], Duda et al. [2001]. Most statistical models have few parameters compared to model-free or nonparametric inference methods.

Given sufficient numbers of parameters and data samples, model-free methods are expected to achieve the best possible performance for any learning task given to them. For instance, given enough training data, optimal decision rules for discriminant functions can be arbitrarily well approximated by consistent nonparametric estimators such as parzen windows, nearest neighbor rules, projection pursuit methods, multilayer ANNs, and classification and regression trees. But, model-free methods are slow to converge, due to excessively large training sets necessary to reduce estimation variances. Model-free architectures may also involve infinitely many parameters, which is an impossible estimation task Geman et al. [1992].

Because convergence speeds for model-free methods are limited by training set size, convergence rates cannot be increased by parallel architectures and fast hardware. The only way to improve convergence speeds is to control the estimation variance. But, controlling the estimation variance requires the use of model-based architectures, which may increase the modeling bias Geman et al. [1992].

### 2.3.1 Essence of the Bias/Variance Dilemma

The essence of the bias/variance dilemma can be summarized as follows. Model-free architectures based on insufficient data samples are unreliable and have slow convergence speeds. However, model-based architectures based on insufficient representations are also unreliable. Except, model-based architectures with adequate representations are both reliable and have reasonable convergence speeds. Even so, model-based architectures with adequate representations are difficult to identify.

All of the above problems indicate that learning unknown functions from data involves obfuscated problems, which remain to be identified.

## 2.4 Prewiring of Important Generalizations

The limitations imposed by the bias and variance dilemma led Geman et al. [1992] to argue that learning complex tasks from training data is essentially impossible without some a priori introduction of carefully designed biases into a learning machine's architecture. They also argue that the identification and

exploitation of proper biases or generalizations are the more fundamental and difficult research issues in neural network modeling applications.

Geman et al. [1992] also suggest that the essential challenges in neural modeling applications are about representation, rather than learning per se. Because most interesting problems tend to be problems of extrapolation, the only way to avoid having to densely cover an input space with training points is to prewire the important generalizations, some of which can be achieved through proper data representations. However, the identification of proper models for complex, statistical inference tasks is a difficult problem. Consequently, any model-based scheme for a complex inference problem may be incorrect, that is, highly biased.

This paper will demonstrate how the identification and exploitation of proper biases or generalizations enables effective designs of learning machine architectures. The paper will address the matter of effective statistical representations for probabilistic, binary, linear classification systems. For the problem of learning decision boundaries, an important form of proper data representations involves the identification and exploitation of pattern vectors which exhibit sufficient class separability, i.e., a negligible overlap exists between data distributions. In general, *the design of effective feature vectors which exhibit sufficient class separability is the most fundamental and difficult problem in the overall design of a statistical pattern recognition system* Fukunaga [1990], Duda et al. [2001].

The essential problem of prewiring the significant statistical representations into a learning machine's architecture remains ill-defined. What does it really mean to introduce a carefully designed bias into a learning machine's architecture? How do we identify the important generalizations for a given problem? How should these generalizations be pre-wired? How does the introduction of a proper bias involve the use of model-based estimation? This paper will consider all of these problems in terms of the fundamental modeling question posed next.

## 2.5 The System Representation Problem

Effective designs of learning machine architectures involve an underlying system modeling problem Naylor and Sell [1971], Gershenfeld [1999]. In particular, effective designs of model-based statistical architectures involve the formulation of a mathematical system which simulates essential stochastic behavior and models key aspects of a real statistical system. Accordingly, *statistical model formulation for learning machine architectures involves the development of a mathematically tractable statistical model that provides a useful representation of a statistical decision system.*

In general terms, a system is an interconnected set of elements which are coherently organized in a manner that achieves a useful function or purpose Meadows [2008]. *This implies that encoding relevant aspects of statistical decision systems within learning machine architectures involves effective interconnections between suitable sets of coherently organized data points.* The learning machine architecture examined in this paper will provide substantial examples of effective interconnections between suitable sets of coherently organized training



points.

## 2.6 Suitable Representations for Learning Machine Architectures

The matter of identifying suitable representations for learning machine architectures is extremely important. Indeed, for many scientific problems, there is a natural and elegant way to represent the solution. For example, each of the well-known special functions, e.g., Legendre polynomials, Bessel functions, Fourier series, Fourier integrals, etc., have the common motivation of being most appropriate for certain problems, and quite unsuitable for others, where each special function represents the relevant aspects of a physical system Keener [2000].

Yet, most machine learning methods attempt to approximate unknown functions with methods that assume no sort of representation, e.g., nonparametric inference methods Geman et al. [1992], Cherkassky and Mulier [1998], Duda et al. [2001], Hastie et al. [2001], Haykin [2009], or assume representations that are tentative and ill-defined, e.g., indefinite interpolations of SVM margin hyperplanes in unknown, high dimensional spaces Boser et al. [1992], Cortes and Vapnik [1995]. Likewise, consider the notion of the asymptotic convergence of a learning machine architecture to some unknown function. Can we picture what this actually means?

## 2.7 Tractable Statistical Models for SVM Architectures

Tangible representations provide objects and forms which can be seen and imagined, along with a perspective for seeing and imagining them Hillman [2012]. This paper will develop substantial geometric architectures for linear SVMs, which are based on a mathematically tractable statistical model that can be depicted and understood in two and three-dimensional vector spaces, and fully comprehended in higher dimensions. The statistical model represents the relevant aspects of a probabilistic, binary linear classification system. The corresponding geometric architecture is determined by correlated, primal and dual, algebraic systems of locus equations of interconnected sets of primal and dual principal eigenaxis components, all of which jointly delineate a linear decision boundary that is bounded by bilaterally symmetrical borders.

The locations of the primal and dual principal eigenaxis components are determined by the geometric and statistical properties of a training data collection. Because a geometric locus of points represents a mathematically tractable geometric curve or surface, the learning machine architecture is completely illustrated in two and three-dimensional vector spaces, and is readily envisioned in higher dimensional vector spaces. The statistical model involves a dual statistical eigenlocus of principal eigenaxis components formed by eigen-scaled extreme data points, all of which encode an eigen-transformed principal location of large covariance. Extreme data points are located in regions of large covariance between two overlapping or non-overlapping data distributions. The term dual statistical eigenlocus is used to describe an eigenlocus, i.e., a characteristic

curve, of principal eigenaxis components which encodes essential geometric underpinnings and statistical machinery for a statistical decision system. A dual statistical eigenlocus is also referred to as a strong dual principal eigenlocus.

## 2.8 A Tractable Dual Locus of Eigen-scaled Data Points

The sections that follow will motivate and develop a learning machine architecture that is based on a natural representation of second-order statistical decision systems. The geometric architecture is based on a dual statistical eigenlocus of principal eigenaxis components which encodes likelihoods of extreme data points. The statistical model provides an elegant solution to difficult interrelated problems that include the bias/variance dilemma, capacity control, and overfitting or underfitting training data. The analysis begins with normal vector directions for linear decision boundaries.

### 2.8.1 Partially Specified Principal Eigenaxes

When an optimum decision boundary is a linear curve or surface, the vector direction deemed most significant is perpendicular to some separating line, plane, or hyperplane. If this direction can be specified in some manner, then no other vector directions contribute useful information for linear discrimination Cooper [1962]. This paper will demonstrate that normal vector directions determine *partially specified principal eigenaxes*, which provide necessary, but insufficient, information for linear decision boundary estimates.

### 2.8.2 Properly Specified Normal Eigenaxes

This paper will show that the most significant vector direction for linear discrimination is specified by the geometric locus of a principal or major eigenaxis, which will be referred to as a normal eigenaxis. It will be shown that the magnitude of a normal eigenaxis contains essential information for effective linear partitioning of pattern vector spaces. Thereby, it will be shown that direction alone is insufficient for describing separating lines, planes, or hyperplanes.

This paper will develop a class of mathematically tractable learning machine architectures that is based on a dual statistical eigenlocus of normal eigenaxis components, all of which encode principal magnitudes and principal directions for linear decision boundary estimates. The paper will extend the fundamental ideas behind the general notion of a geometric locus to develop symmetrical algebraic systems of primal and dual normal eigenlocus equations that jointly specify a normal eigenlocus of eigen-scaled extreme data points. The term strong dual normal eigenlocus refers to a dual statistical eigenlocus of normal eigenaxis components. Extreme data points are innermost data points of large covariance which are located between overlapping or non-overlapping data distributions. The analyses presented in this paper will demonstrate how correlated algebraic systems of primal and dual normal eigenlocus equations provide an estimate of an unknown normal eigenaxis of an unknown linear decision boundary. Because

a strong dual normal eigenlocus is based on graphs or geometric loci of equations, a strong dual normal eigenlocus is mathematically tractable.

To motivate the development of a strong dual normal eigenlocus of eigen-scaled extreme data points, Section 3 will consider the fundamental limitations of classical geometric locus methods applied to collections of training data. These limitations will be described in terms of the geometric locus dilemma for statistical learning machines, a terminology inspired by Geman et al. [1992]. Existing locus methods are outlined next.

## 2.9 The Graph or Locus of an Equation

The graph or locus of an equation is the locus (place) of all points whose coordinates are solutions of the equation. Any point whose coordinates are solutions of a locus equation is on the geometric locus of the equation. Any given point on a geometric locus possesses a geometric property which is common to all points on the locus, and *no other points* Nichols [1893], Tanner and Allen [1898], Whitehead [1911], Eisenhart [1939]. For example, consider the geometric locus of a circle. A circle is a locus of points  $(x, y)$ , all of which are at the same distance, the radius  $r$ , from a fixed point  $(x_0, y_0)$ , the center. The algebraic equation for the geometric locus of a circle in Cartesian coordinates is:

$$(x - x_0)^2 + (y - y_0)^2 = r^2. \quad (1)$$

For any given center  $(x_0, y_0)$  and radius  $r$ , only those coordinates  $(x, y)$  that satisfy Eq. (1) contribute to the geometric locus of the specified circle Eisenhart [1939].

The identification of the geometric property of a locus of points is a central problem in coordinate geometry. The inverse problem finds the algebraic form of an equation, whose solution gives the coordinates of all of the points on a locus which has been defined geometrically. Geometric figures are defined in two ways: (1) as a figure with certain known properties, and (2) as the path of a point which moves under known conditions Nichols [1893], Tanner and Allen [1898].

## 2.10 Methods for Solving Locus Problems

Methods for solving geometric locus problems hinge on the identification of algebraic and geometric correlations for a given locus of points. Geometric correlations between a set of points which lie on a definite curve or surface, correspond to geometric and algebraic constraints that are satisfied by the coordinates of any point on a given locus Nichols [1893], Tanner and Allen [1898], Whitehead [1911], Eisenhart [1939].

Finding the algebraic form of an equation for a given geometric figure or locus is often a difficult problem. However, because some type of relationships exist between a geometric locus and its algebraic form, careful examination into the point and coordinate relationships specified by the algebraic form of a locus

equation may yield additional insight into the uniform property exhibited by a geometric locus of points Tanner and Allen [1898].

## 2.11 Changing the Loci of Reference Axes

The algebraic form of a locus equation hinges on both the geometric property and the frame of reference (coordinate system) of the locus. Thereby, changing the position of the coordinate axes changes both (1) the algebraic form of the locus that references the new axes and (2) the coordinates of any point on the locus. It follows that the equation of a locus and the identification of the geometric property of the locus can be greatly simplified by changing the position of the axes to which the locus of points is referenced Nichols [1893], Tanner and Allen [1898], Eisenhart [1939].

## 2.12 Geometric Locus of a Straight Line

The geometric locus of every equation of the first degree is a straight line. Only two geometric conditions are deemed necessary to determine the equation of a particular line. Either a line should pass through two given points, or should pass through a given point and have a given slope. Standard equations of a straight line include the point-slope, slope-intercept, two-point, intercept, and normal forms.

The general equation of the first degree in two coordinate variables  $x$  and  $y$  has the form:

$$Ax + By + C = 0,$$

where  $A$ ,  $B$ ,  $C$  are constants which may have any real values, subject to the restriction that  $A$  and  $B$  cannot both be zero Nichols [1893], Tanner and Allen [1898], Eisenhart [1939].

Excluding the point, a straight line appears to be the simplest type of geometric locus. Nonetheless, the uniform geometric property of a straight line remains undefined. This paper will soon identify several, correlated uniform properties exhibited by all of the points on a linear locus.

The next section will outline the algebraic, geometric, and statistical essence of the geometric locus dilemma for statistical learning machines. The essence of the dilemma will be defined for support vector learning machines, and a method will be outlined that resolves the dilemma for linear and polynomial kernel SVMs. By way of motivation, simulation studies will be presented which demonstrate that linear kernel SVMs learn Bayes' decision boundaries for training data drawn from overlapping Gaussian distributions.

## 3 The Geometric Locus Dilemma

An insoluble dilemma in the design and development of learning machine architectures will now be identified. The dilemma underlies interrelated problems

that include the bias/variance dilemma, capacity control, and overfitting or underfitting training data. The underlying issue involves determining an effective fit of an  $N$ -dimensional set of  $d$ -dimensional random data points to geometric loci in  $d$ -dimensional Cartesian space. More specifically, the problem involves defining suitable fits for given collections of  $N \times d$  random vector coordinates to algebraic equations of prespecified (explicit) geometric loci that reference arbitrary Cartesian coordinate systems.

A classical locus of points is an explicit, and thus fixed, geometric configuration of vectors, whose Cartesian coordinate locations are determined by, and therefore satisfy, an algebraic equation. Curves or surfaces of classical locus equations are determined by properties of geometric loci with respect to coordinate axes of arbitrary Cartesian coordinate systems. Thereby, an algebraic equation of a classical locus of points generates an explicit point, curve, or surface, in an arbitrarily specified Cartesian space. It follows that any point on a classical geometric locus naturally exhibits the uniform property of the locus. Indeed, any point on a classical geometric locus satisfies the uniform property of the given locus in an innate manner. It will now be argued that fitting collections of training data to classical geometric locus equations involves an impossible estimation task.

### 3.1 An Impossible Estimation Task

Consider fitting a collection of training data to some classical or standard locus equation(s). It follows that any given second-order curve or surface must pass through any training points on the specified locus. Therefore, any training point whose coordinate locations do not satisfy the given locus equation and correlated geometric property of the specified locus is simply not on the given curve or surface. Likewise, training points that are not on a given curve or surface do not contribute to the locus of a specified curve or surface.

Given the correlated algebraic and geometric constraints on a classical locus of points, it follows that any attempt to *fit* an  $N$ -dimensional set of  $d$ -dimensional random data points to the equation(s) of a classical geometric locus, involves the unfeasible problem of determining an *effective constellation* of an  $(N - M) \times d$  subset of  $N \times d$  random vector coordinates that (1) inherently satisfy prespecified, fixed magnitude (length) constraints on each of the respective  $d$  Cartesian coordinate axes, and thereby (2) generate explicit, fixed points, curves, or surfaces in  $\mathbb{R}^d$ . Such an estimation process is clearly unfeasible. It follows that fitting collections of random data points to classical locus equations is an impossible estimation task. The essence of the geometric locus dilemma for support vector learning machines is defined next.

### 3.2 SVMs and the Geometric Locus Dilemma

So far, it has been argued that the insoluble aspect of the geometric locus dilemma concerns finding suitable fits for collections of random vector coordinates to algebraic equations of partially configured geometric loci that reference

arbitrary Cartesian coordinate systems. It has also been argued that such estimation processes are impracticable methods that involve impossible estimation tasks. The next section will argue that SVM capacity control involves an impossible estimation task.

### 3.2.1 Linear Interpolation Using Random Slack Variables

SVM methods are based on the idea of specifying a pair of maximally separated linear curves or surfaces that interpolate two sets of data points Cortes and Vapnik [1995], Bennett and Campbell [2000], Cristianini and Shawe-Taylor [2000], Hastie et al. [2001], Scholkopf and Smola [2002]. Given two non-overlapping sets of data points, linear SVM finds a pair of maximally separated linear decision borders, such that the decision borders pass through data points called support vectors. For example, in Section 2, Fig. 2 illustrates how SVM decision borders interpolate two sets of data points, where the green decision borders exhibit the largest geometric margin.

Identifying interpolation methods that provide effective fits of separating lines, planes, or hyperplanes involves the long standing problem of fitting linear decision boundaries to overlapping sets of data points Cover [1965]. Soft margin linear SVM is said to resolve this problem by means of non-negative random slack variables  $\xi_i \geq 0$ , each of which allows a correlated data point  $\mathbf{x}_i$ , that lies between or beyond a pair of linear decision borders, to satisfy a linear border. Nonlinear kernel SVMs also employ non-negative random slack variables, each of which allows a transformed, correlated data point to satisfy a hyperplane decision border in some higher dimensional feature space Cortes and Vapnik [1995], Bennett and Campbell [2000], Cristianini and Shawe-Taylor [2000], Hastie et al. [2001], Scholkopf and Smola [2002].

*This implies that non-negative random slack variables, for both linear and nonlinear kernel SVMs, encode effective distances of data points from unknown linear curves or surfaces.* Clearly, this is an impossible estimation task. It follows that computing effective values for  $l$  non-negative random slack variables  $\{\xi_i | \xi_i \geq 0\}_{i=1}^l$  is an impossible estimation task. Figure 3 depicts the insoluble aspect of the geometric locus dilemma for linear kernel SVMs.

### Linear Interpolations of Data Points Using Random Slack Variables $\xi_i \geq 0$

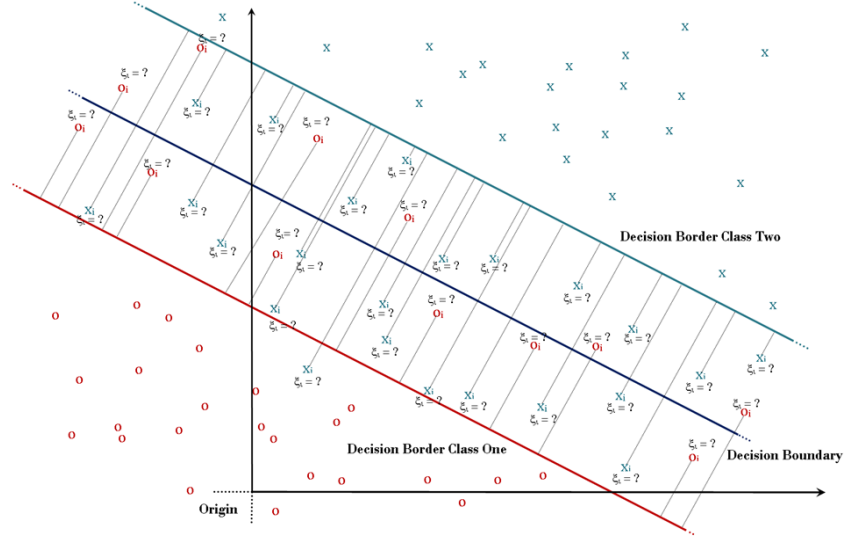


Figure 3: Illustration of the geometric locus dilemma for linear kernel SVMs:  $l$  non-negative random slack variables  $\{\xi_i | \xi_i \geq 0\}_{i=1}^l$  must be estimated for effective linear interpolations of the  $l - k$  blue  $\{\mathbf{x}_i\}_{i=1}^{l-k}$  and the  $l - k + 1$  red  $\{\mathbf{o}_i\}_{i=l-k+1}^l$  overlapping data points. Computing effective values for these  $l$  random slack variables is an impossible estimation task.

### 3.3 Ill-defined SVM Architectures

SVM architectures and regularization parameters are largely ill-defined. For example, the polynomial degree, kernel width, and regularization parameters of nonlinear kernel SVMs, and the regularization parameters of soft margin linear kernel SVMs, are mostly determined by trial and error Byun and Lee [2002], Eitrich and Lang [2006], Liang et al. [2011]. Likewise, the total citation count for the 1998 article titled *A Tutorial on Support Vector Machines for Pattern Recognition* Burges [1998] currently exceeds 15,000.

This paper will demonstrate that resolving the geometric locus dilemma for linear and polynomial kernel SVMs involves two correlated and fundamental problems that involve graphs or loci of properly specified locus equations. This paper will show that effective design of linear kernel SVM architectures are based on (1) a suitable set of geometric loci that provide the basis of a statistical decision system, and (2) a properly specified algebraic system of locus equations for a given statistical decision system. This paper will also establish that the fundamental geometric locus of interest for linear kernel SVM architectures is the *locus of a principal eigenaxis*. Polynomial kernel SVMs will be extensively

examined in an upcoming paper. At this time, the paper will motivate taking an extensive look under the hood of linear kernel SVMs.

### 3.4 How Does Linear Kernel SVM Learn from Data?

Linear kernel SVMs have been applied to training data drawn from overlapping Gaussian distributions Reeves [2007], Reeves [2009]. The results obtained from these simulation studies indicate that linear kernel SVM learns Bayes' decision boundaries for highly overlapping Gaussian data sets. Bayes' discriminant functions are the gold standard for linear discrimination tasks. Bayes' classification error rate is the best error rate that can be achieved by any classifier Fukunaga [1990], Duda et al. [2001]. Results from two of the simulation studies are outlined below.

#### 3.4.1 Classification Example One

Gaussian data set one has the covariance matrix:

$$\Sigma_1 = \Sigma_2 = \begin{pmatrix} 0.5 & 0 \\ 0 & 2 \end{pmatrix},$$

and the mean vectors  $\mu_1 = (3, 0.5)^T$  and  $\mu_2 = (3, -0.5)^T$ . The Bayes' decision boundary

$$x_2 = 0,$$

which is depicted in Fig. 4, enforces the Bayes' error rate of 36.5%.

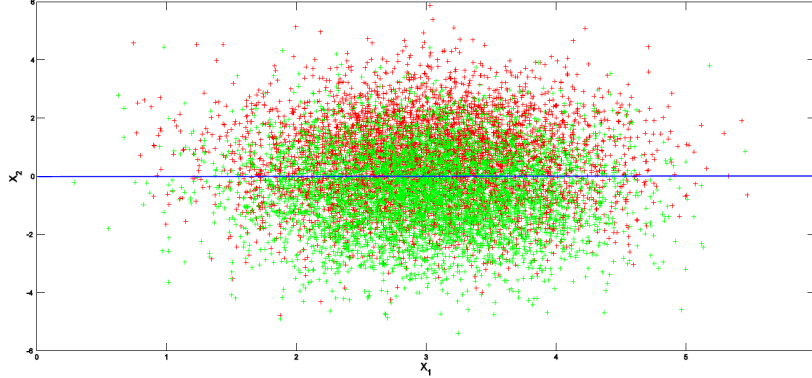


Figure 4: The Bayes' decision boundary for the overlapping Gaussian data sets of classification example one.

Figure 5 illustrates that linear kernel SVM learns the Bayes' decision boundary for the overlapping Gaussian data in classification example one. Linear kernel SVM uses 596 support vectors (99% of the training data) to learn the Bayes' decision boundary depicted in Fig. 4.



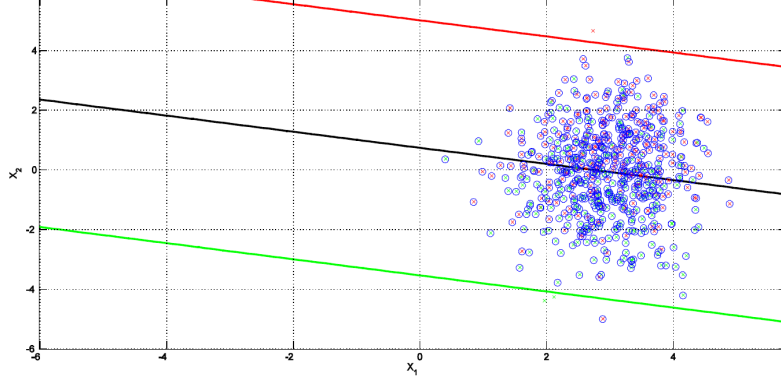


Figure 5: Linear kernel SVM learns the Bayes' decision boundary for the overlapping Gaussian data sets of classification example one. Each support vector is enclosed in a blue circle.

### 3.4.2 Classification Example Two

Gaussian data set two has the covariance matrix:

$$\Sigma_1 = \Sigma_2 = \begin{pmatrix} 0.95 & 0.45 \\ 0.45 & 0.35 \end{pmatrix},$$

and the mean vectors  $\mu_1 = (3, 0.25)^T$  and  $\mu_2 = (3, -0.25)^T$ . The Bayes' decision boundary

$$x_2 = 0.47x_1 - 1.42,$$

which is depicted in Fig. 6, enforces the Bayes' error rate of 25%.

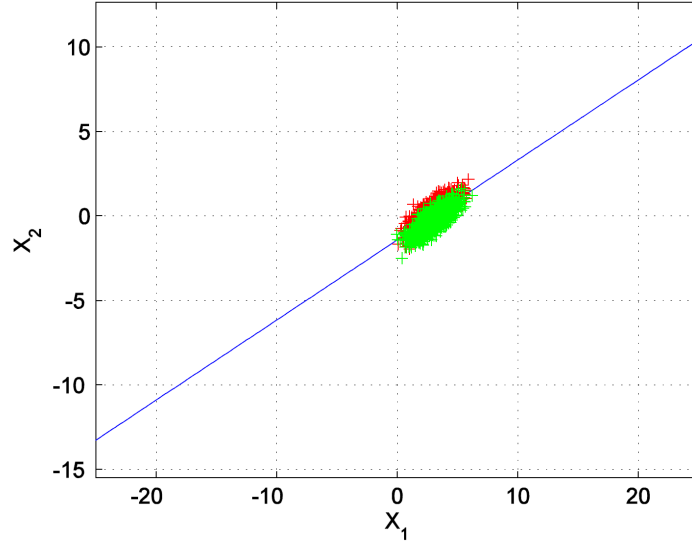


Figure 6: The Bayes' decision boundary for the overlapping Gaussian data sets of classification example two.

Figure 7 illustrates that linear kernel SVM also learns the Bayes' decision boundary for the overlapping Gaussian data in classification example two. Linear kernel SVM uses 547 support vectors (91% of the training data) to learn the Bayes' decision boundary depicted in Fig. 6.

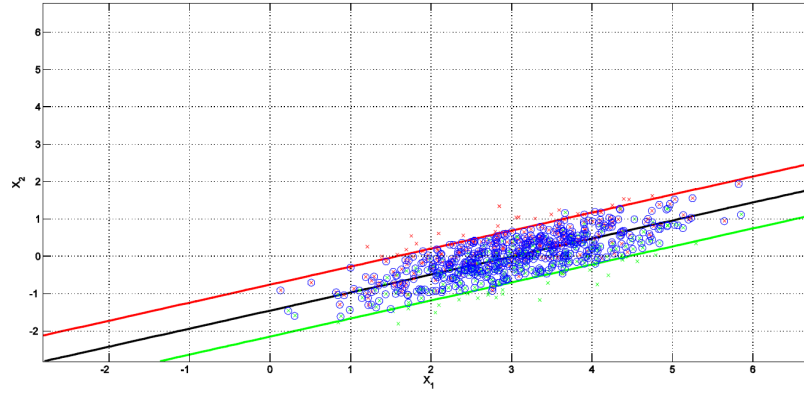


Figure 7: Linear kernel SVM learns the Bayes' decision boundary for the overlapping Gaussian data sets in classification example two. Each support vector is enclosed in a blue circle.

Consideration of the above simulation studies *motivates an extensive investigation into what is really happening under the hood of linear kernel SVMs*. How does linear SVM actually learn Bayes' linear decision boundaries for overlapping Gaussian data distributions? How does the geometric architecture of linear kernel SVM encode Bayes' likelihood ratio? How do we describe linear kernel SVM architectures? What types of geometric underpinnings make up the statistical machinery of linear kernel SVM architectures?

All of these questions will be resolved in the sections that follow. The next section of the paper will begin the process of taking a comprehensive look under the hood of linear kernel SVMs. Section 4 will develop an elegant principal eigen-coordinate system that describes all forms of linear loci.

## 4 An Elegant Principal Eigen-coordinate System for Linear Loci

What types of geometric underpinnings and statistical machinery are encoded within linear kernel SVM architectures? How does linear SVM learn optimal linear decision boundaries for overlapping data distributions? All of these questions will be answered by describing the linear SVM method in a geometric locus framework. Moreover, answers to these question will identify an estimation method that resolves the geometric locus dilemma for linear SVMs. It will be shown that learning linear decision boundaries involves strong duality relationships, between a statistical eigenlocus of principal eigenaxis components and its algebraic forms, in primal and dual, correlated inner product, i.e., Hilbert, spaces. Thereby, a computer-implemented method will be formulated that generates regularized, data-driven geometric architectures which encode Bayes' likelihood ratio for common covariance data and a robust likelihood ratio for all other data distributions.

Section 4 begins by introducing and developing a locus equation of the principal eigenaxis of a straight line. Overviews of the equations of a straight line can be found in Nichols [1893], Tanner and Allen [1898], Eisenhart [1939], and Davis [1973]. The vector equation of a straight line, which is outlined in Davis [1973], is the hinge point and central principle for the chain of arguments on linear decision boundary estimates that follow.

### 4.1 Locus Equations of a Normal Eigenaxis

At this time, a few remarks on notation are necessary. Strictly speaking, a vector  $\mathbf{x}$  is a directed straight line segment that emanates from a chosen point  $P_0$ , termed the origin, such that the endpoint of the directed straight line segment, termed the tip, defines a real specific point  $P$ . Thereby, a point is an inherent part of a vector, such that correlated points  $P_{\mathbf{x}}$  and vectors  $\mathbf{x}$  both describe the same ordered pair of real numbers in the real Euclidean plane or the same ordered  $d$ -tuple of real numbers in real Euclidean space. Given that correlated points  $P_{\mathbf{x}}$  and vectors  $\mathbf{x}$  specify equivalent ordered pairs or  $d$ -tuples

of real numbers, correlated points  $P_{\mathbf{x}}$  and vectors  $\mathbf{x}$  have common geometric representations as points in  $\mathbb{R}^2$  or  $\mathbb{R}^d$ . Depending on the geometric context, an ordered pair of real numbers or an ordered  $d$ -tuple of real numbers will be referred to as either a point or a vector. The analysis that follows will denote points and vectors by  $\mathbf{x}$ .

A locus equation is now introduced that describes lines simply in terms of algebraic correlations between the points on a line. Moreover, the locus equation contains no constants or parameters. It will be argued that one of the points on any given line determines the principal eigenaxis of the line. The principal eigenaxis of a linear locus will be called a *normal eigenaxis*. Several locus equations of a normal eigenaxis will be developed, all of which describe geometric loci of lines, planes, and hyperplanes. All of these algebraic expressions will be used to identify the correlated uniform properties exhibited by any point on a linear locus.

The development of normal eigenaxis locus equations and the identification of the correlated uniform properties exhibited by any point on a linear locus, will lead to far-reaching consequences for the matter of linear decision boundary estimates. The locus equations of a normal eigenaxis and the invariant geometric properties of a normal eigenaxis will be used to motivate and develop symmetric primal and dual algebraic systems of strong dual normal eigenlocus equations, which will be demonstrated to jointly specify robust and optimal estimates of separating lines, planes, and hyperplanes for a large number of data distributions, including homogeneous data distributions.

The paper will now derive the primary locus equation of a normal eigenaxis that describes geometric loci of lines, planes, and hyperplanes. By way of introduction, Fig. 8 depicts the geometric locus of a line in the real Euclidean plane  $\mathbb{R}^2$ .

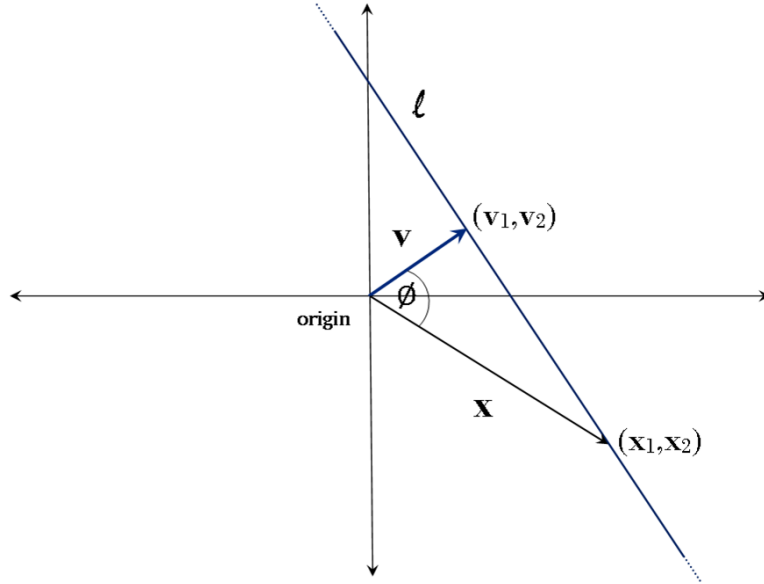


Figure 8: Illustration of an elegant principal eigen-coordinate system for lines that is readily generalized to planes and hyperplanes. Any vector  $\mathbf{x}_i = (x_{1i}, x_{2i})^T$  whose tip  $(x_{1i}, x_{2i})$  is on the line  $l$  explicitly and exclusively references the vector  $\mathbf{v} = (v_1, v_2)^T$ , which is shown to be the principal, i.e., normal, eigenaxis of the line  $l$ .

## 4.2 The Normal Eigenaxis of a Linear Locus

Let  $\mathbf{v} \triangleq (v_1, v_2)^T$  be a fixed vector in the real Euclidean plane and consider the line  $l$  at the tip of  $\mathbf{v}$  that is perpendicular to  $\mathbf{v}$ . To wit, the vector  $\mathbf{v}$  is a point on the line  $l$ . Therefore, the coordinates  $(v_1, v_2)$  of  $\mathbf{v}$  delineate and satisfy  $l$ . In addition, consider an arbitrary vector  $\mathbf{x} \triangleq (x_1, x_2)^T$  whose tip is also on the line  $l$ . Thereby, the coordinates  $(x_1, x_2)$  of the point  $\mathbf{x}$  also delineate and satisfy  $l$ . Finally, let  $\phi$  be the acute angle between the vectors  $\mathbf{v}$  and  $\mathbf{x}$ , satisfying  $0 \leq \phi \leq \pi/2$  and the algebraic relationship  $\cos \phi = \frac{\|\mathbf{v}\|}{\|\mathbf{x}\|}$ .

Using all of the above assumptions, it follows that the locus of points  $(x_1, x_2)$  on the line  $l$  is described by the functional locus equation:

$$\mathbf{x}^T \mathbf{v} = \|\mathbf{x}\| \|\mathbf{v}\| \cos \phi, \quad (2)$$

which is the vector equation of a line Davis [1973]. It will shortly be demonstrated that the vector  $\mathbf{v}$  is the principal eigenaxis of the line  $l$ .

### 4.3 Fundamental Equation of a Linear Locus

Take any fixed vector  $\mathbf{v}$ , and consider the line  $l$  that is described by Eq. (2), where the axis of  $\mathbf{v}$  is perpendicular to the specified line  $l$  and the tip of  $\mathbf{v}$  is on  $l$ . Given that any vector  $\mathbf{x}_i$  with its tip on the given line  $l$  satisfies the algebraic relationship:

$$\|\mathbf{x}_i\| \cos \phi_i = \|\mathbf{v}\|,$$

with the fixed vector  $\mathbf{v}$ , it follows that the geometric locus of a line  $l$  is also described by the locus equation:

$$\mathbf{x}^T \mathbf{v} = \|\mathbf{v}\|^2. \quad (3)$$

Equations (2) and (3) are readily generalized to planes  $p$  and hyperplanes  $h$  in  $\mathbb{R}^d$  by letting  $\mathbf{v} \triangleq (v_1, v_2, \dots, v_d)^T$  and  $\mathbf{x} = (x_1, x_2, \dots, x_d)^T$ . Given that Eq. (3) contains no constants or parameters, it follows that Eq. (3) is the fundamental equation of a linear locus.

Assuming that  $\|\mathbf{v}\| \neq 0$ , Eq. (3) can also be written as:

$$\frac{\mathbf{x}^T \mathbf{v}}{\|\mathbf{v}\|} = \|\mathbf{v}\|. \quad (4)$$

The axis  $\mathbf{v}/\|\mathbf{v}\|$  has length 1 and points in the direction of the vector  $\mathbf{v}$ , such that  $\|\mathbf{v}\|$  is the distance of a specified line  $l$ , plane  $p$ , or hyperplane  $h$  to the origin. Using Eq. (4), it follows that the distance  $\Delta$  of a line, plane, or hyperplane from the origin is specified by the magnitude  $\|\mathbf{v}\|$  of the axis  $\mathbf{v}$ .

It is claimed that the vector  $\mathbf{v}$  provides an exclusive, intrinsic reference axis for a linear locus of points. An intrinsic axis which coincides as an exclusive, fixed reference axis for coordinates, delineates curves or surfaces that are mirror images of each other. Such intrinsic reference axes are principal eigenaxes of conic sections and quadratic surfaces Hewson [2009].

### 4.4 Major Intrinsic Axes of Second-order Loci

All of the second-order geometric loci are characterized by one or more intrinsic axes which are represented by the eigenvectors of a real symmetric matrix associated with a quadratic form. This paper claims that a major intrinsic axis, which may coincide as an exclusive, fixed reference axis, is an inherent part of any second-order geometric locus. Examples of major intrinsic axes include the major axes, i.e., principal eigenaxes, of ellipses, parabolas, and hyperbolas. In general, geometric shapes and orientations of conic sections and quadratic surfaces are described by eigenvalues and eigenaxes Hewson [2009].

Recall that the algebraic form of a locus equation hinges on both the geometric property and the frame of reference of the locus. Because coordinate versions of geometric loci reference Cartesian coordinate systems, the positions of axes of coordinates to which a given locus is referenced are arbitrary Nichols [1893], Tanner and Allen [1898], Eisenhart [1939]. However, this paper claims that the positions of the major axes of the second-order geometric loci are not arbitrary.

It can be argued that the locus of a major axis is characteristic of a particular second-order locus of points. It can also be demonstrated that a principal eigenaxis offers an elegant principal eigen-coordinate system for a conic section or quadratic surface. It will shortly be demonstrated that the locus of the major intrinsic axis of a linear locus offers an elegant principal eigen-coordinate system that is characteristic of a specific locus of points.

#### 4.5 The Principal Eigenaxis of a Linear Locus

Figure 8 shows how the geometric configuration of a fixed vector  $\mathbf{v}$  determines the geometric configuration of a linear locus  $l$ . It will now be demonstrated that the axis  $\mathbf{v}$  denoted in Eqs (2), (3), and (4) is the principal eigenaxis of linear loci.

All of the major axes of conic sections and quadratic surfaces are major intrinsic axes which may also coincide as exclusive, fixed reference axes Nichols [1893], Tanner and Allen [1898], Eisenhart [1939]. In order to demonstrate that  $\mathbf{v}$  is the principal eigenaxis of linear loci, it must be shown that  $\mathbf{v}$  is a major intrinsic axis which is also a reference axis. It will first be argued that  $\mathbf{v}$  is a major intrinsic axis for a linear locus of points.

Using the definitions of Eqs (2), (3), or (4), the axis  $\mathbf{v}$  is a major intrinsic axis because all of the points  $\mathbf{x}$  on a linear locus satisfy identical algebraic and geometric constraints related to the locus of  $\mathbf{v}$  that are inherently specified by Eqs (2), (3), and (4). Therefore,  $\mathbf{v}$  is a major intrinsic axis of a linear locus. The uniform algebraic and geometric constraints satisfied by all of the points on a linear locus determine the uniform properties exhibited by each point on the linear locus.

Again using the definitions of Eqs (2), (3), or (4), the axis  $\mathbf{v}$  is an exclusive reference axis because the uniform properties possessed by all of the points  $\mathbf{x}$  on a linear locus are defined with respect to the axis of  $\mathbf{v}$ . Therefore, all of the points  $\mathbf{x}$  on a given line, plane, or hyperplane, explicitly and exclusively reference the major intrinsic axis  $\mathbf{v}$  of the linear locus. It is concluded that the vector  $\mathbf{v}$  provides an exclusive, fixed reference axis for a linear locus.

In conclusion, it has been demonstrated that the vector  $\mathbf{v}$  is a major intrinsic axis that coincides as an exclusive, fixed reference axis for a linear locus. It follows that  $\mathbf{v}$  is the major axis of linear loci. It is concluded that the vector  $\mathbf{v}$  denoted in Eqs (2), (3), and (4) is the principal eigenaxis of linear curves and plane or hyperplane surfaces in  $\mathbb{R}^d$ .

It will now be argued that the locus of a principal eigenaxis  $\mathbf{v}$  is unique. Take any given line  $l$ , plane  $p$ , or hyperplane  $h$ . Using the definitions of Eqs (2), (3), or (4), it follows that the given line  $l$ , plane  $p$ , or hyperplane  $h$  is perpendicular to the principal eigenaxis  $\mathbf{v}$  of the specified line  $l$ , plane  $p$ , or hyperplane  $h$ , at the tip of the principal eigenaxis  $\mathbf{v}$ . Moreover, the line  $l$ , plane  $p$ , or hyperplane  $h$  is perpendicular to only one major axis  $\mathbf{v}$ , at the tip of that major axis  $\mathbf{v}$ , which is uniquely specified by the locus of  $\mathbf{v}$ . It is concluded that the locus of a principal eigenaxis is unique for any given linear locus of points.

The vector  $\mathbf{v}$  will be referred to as the normal eigenaxis of linear curves and surfaces. Equation (4) will now be used to develop a coordinate form locus equation, which will be used to identify a uniform property which is exhibited by any point on a linear curve or surface.

#### 4.6 Coordinate Form Locus Equation of a Unit Normal Eigenaxis

Given Eq. (4), it follows that any line  $l$  in the Euclidean plane  $\mathbb{R}^2$  and any plane  $p$  or hyperplane  $h$  in Euclidean space  $\mathbb{R}^d$  is described by the locus equation

$$\mathbf{x}^T \mathbf{u}_{N_e} = \Delta, \quad (5)$$

where  $\mathbf{u}_{N_e}$  is a unit length normal eigenaxis that is perpendicular to  $l$ ,  $p$ , or  $h$ , and  $\Delta$  denotes the distance of  $l$ ,  $p$ , or  $h$  to the origin. The unit eigenvector  $\mathbf{u}_{N_e}$  specifies the direction of a normal eigenaxis of a linear curve or surface, while the distance  $\Delta$  of a line, plane, or hyperplane from the origin is specified by the magnitude  $\|\mathbf{v}\|$  of its normal eigenaxis  $\mathbf{v}$ .

Now express  $\mathbf{u}_{N_e}$  in terms of standard orthonormal basis vectors

$$\{\mathbf{e}_1 = (1, 0, \dots, 0), \dots, \mathbf{e}_d = (0, 0, \dots, 1)\}$$

so that

$$\mathbf{u}_{N_e} = \cos \alpha_1 \mathbf{e}_1 + \cos \alpha_2 \mathbf{e}_2 + \dots + \cos \alpha_d \mathbf{e}_d,$$

where  $\cos \alpha_i$  are the direction cosines between  $\mathbf{u}_{N_e}$  and  $\mathbf{e}_i$ . Each  $\cos \alpha_i$  is the  $i^{\text{th}}$  component of the unit normal eigenaxis  $\mathbf{u}_{N_e}$  along the coordinate axis  $\mathbf{e}_i$ , where each eigen-scale  $\cos \alpha_i$  is said to be normalized.

Substitution of the expression for  $\mathbf{u}_{N_e}$  into Eq. (5) produces a coordinate form locus equation

$$x_1 \cos \alpha_1 + x_2 \cos \alpha_2 + \dots + x_d \cos \alpha_d = \Delta, \quad (6)$$

which is satisfied by the eigen-transformed coordinates  $(\cos \alpha_1 x_1, \dots, \cos \alpha_d x_d)$  of all of the points  $\mathbf{x}$  on the geometric locus of a line, plane, or hyperplane. Equation (6) is the well known coordinate equation version of a linear locus

$$\cos \alpha_1 x_1 + \cos \alpha_2 x_2 + \dots + \cos \alpha_d x_d = \Delta,$$

which is usually written as

$$Ax_1 + Bx_2 + \dots + Nx_N = P.$$

Equation (6) is now used to define a uniform property which is exhibited by any point on a linear locus.



## 4.7 Uniform Property Exhibited by Points on a Linear Locus

Given Eq. (6), it follows that a line, plane, or hyperplane is a locus of points  $\mathbf{x}$ , all of which possess a set of normalized, eigen-scaled coordinates:

$$\mathbf{x} = (\cos \alpha_1 x_1, \cos \alpha_2 x_2, \dots, \cos \alpha_d x_d)^T,$$

such that the sum of those coordinates equals the distance  $\Delta$  that the line, plane, or hyperplane is from the origin  $(0, 0, \dots, 0)$ :

$$\sum_{i=1}^d \cos \alpha_i x_i = \Delta, \quad (7)$$

where  $x_i$  are point coordinates or vector components, and  $\cos \alpha_i$  are the direction cosines between a unit length normal eigenaxis  $\mathbf{u}_{N_e}$  and the coordinate axes  $\mathbf{e}_i : \{\mathbf{e}_1 = (1, 0, \dots, 0), \dots, \mathbf{e}_d = (0, 0, \dots, 1)\}$ .

It follows that a point  $\mathbf{x}$  is on the geometric locus of a line  $l$ , plane  $p$ , or hyperplane  $h$ , if, and only if, the normalized, eigen-scaled coordinates of  $\mathbf{x}$  satisfy Eq. (7); otherwise, the point  $\mathbf{x}$  is not on the locus of points described by Eqs (2), (3), and (5). Given Eq. (7), it follows that the sum of normalized, eigen-scaled coordinates of any point on a linear locus, equals the magnitude of the normal eigenaxis of the linear locus. It is concluded that all of the points  $\mathbf{x}$  on a linear locus possess a characteristic set of eigen-scaled coordinates, such that the inner product of each vector  $\mathbf{x}$  with  $\mathbf{u}_{N_e}$  satisfies the distance  $\Delta$  of the linear locus from the origin.

So far, the principal eigenaxis of linear curves and surfaces has been identified, along with a correlated uniform property which is exhibited by any point on the lines, planes, or hyperplanes of Eqs (2), (3), and (4). Moreover, Eqs (2) - (7) all indicate that the eigen-coordinate locations of a normal eigenaxis provide a distinctive set of eigenfeatures which effectively characterize all forms of lines, planes, and hyperplanes. To wit, the locus of a normal eigenaxis effectively determines the locus of points on a linear curve or surface. This implies that the important generalizations for a linear locus are encoded within the invariant geometric location of its normal eigenaxis.

The next section will examine how important generalizations and properties for linear loci are encoded within the geometric locus of a normal eigenaxis. It will be demonstrated that a normal eigenaxis is an exclusive, intrinsic reference axis which has a distinctive geometric configuration, a characteristic set of eigen-coordinate loci, and a characteristic eigenenergy, all of which determine a characteristic eigen-signature for any given linear locus of points. The uniform properties which are satisfied by all of the points on a linear locus will be identified. It will also be demonstrated that each uniform property is uniquely determined by the locus of a normal eigenaxis. Later on, these arguments will be extended to include linear decision boundary estimates. The properties of normal eigenaxes are examined next.

## 4.8 Properties of Normal Eigenaxes

Take any line, plane, or hyperplane in  $\mathbb{R}^d$ . Given Eqs (2) or (3) and a particular line, plane, or hyperplane, it follows that a normal eigenaxis  $\mathbf{v}$  exists, such that the tip of  $\mathbf{v}$  is on the specified line, plane, or hyperplane, and the axis of  $\mathbf{v}$  is perpendicular to the line, plane, or hyperplane. Given Eq. (4), it follows that the length  $\|\mathbf{v}\|$  of  $\mathbf{v}$  is determined by the given line, plane, or hyperplane. Given Eq. (6), it follows that the unit normal eigenaxis  $\mathbf{u}_{N_e}$  of the specified linear curve or surface is characterized by a unique set of direction cosines  $\cos \alpha_i$  between  $\mathbf{u}_{N_e}$  and each standard basis vector  $\mathbf{e}_i$ .

Next, take any normal eigenaxis  $\mathbf{v}$  in  $\mathbb{R}^d$ . Given Eqs (2) or (3) and a particular normal eigenaxis  $\mathbf{v}$ , it follows that a line, plane, or hyperplane exists that is perpendicular to  $\mathbf{v}$ , such that the tip of the given normal eigenaxis  $\mathbf{v}$  is on the line, plane, or hyperplane. Given Eq. (4), it follows that the distance of the line, plane, or hyperplane from the origin is specified by the magnitude  $\|\mathbf{v}\|$  of the given normal eigenaxis  $\mathbf{v}$ . It will now be shown that the normal eigenaxis of any given linear locus satisfies the linear locus in terms of its eigenenergy.

### 4.8.1 Characteristic Eigenenergy of a Normal Eigenaxis

Take the normal eigenaxis  $\mathbf{v}$  of any line, plane, or hyperplane in  $\mathbb{R}^d$ . It follows that the normal eigenaxis  $\mathbf{v}$  satisfies Eqs (2), (3), and (4). Given Eqs (2) or (3) and a particular normal eigenaxis  $\mathbf{v}$ , it follows that the normal eigenaxis  $\mathbf{v}$  satisfies the linear locus in terms of its eigenenergy:

$$\mathbf{v}^T \mathbf{v} = \|\mathbf{v}\|^2.$$

Therefore, the normal eigenaxis  $\mathbf{v}$  of any given line, plane, or hyperplane exhibits a characteristic eigenenergy  $\|\mathbf{v}\|^2$ . It follows that the line, plane, or hyperplane delineated by a normal eigenaxis  $\mathbf{v}$  exhibits a characteristic eigen-signature in the form of the characteristic eigenenergy  $\|\mathbf{v}\|^2$  of its normal eigenaxis  $\mathbf{v}$ .

It is concluded that the normal eigenaxis of any given linear locus satisfies the linear locus in terms of its eigenenergy. It is also concluded that the fundamental property of any given normal eigenaxis  $\mathbf{v}$  is its characteristic eigenenergy  $\|\mathbf{v}\|^2$ .

## 4.9 Correlated Uniform Properties Exhibited by Points on a Linear Locus

Take any point  $\mathbf{x}$  on any linear locus. Given Eq. (2) and a specific point  $\mathbf{x}$  on a particular locus, it follows that the length of the component  $\|\mathbf{x}\| \cos \phi$  of the given vector  $\mathbf{x}$  along the normal eigenaxis  $\mathbf{v}$  of the given locus satisfies the length  $\|\mathbf{v}\|$  of  $\mathbf{v}$ :

$$\|\mathbf{x}\| \cos \phi = \|\mathbf{v}\|,$$

where the length  $\|\mathbf{v}\|$  of  $\mathbf{v}$  determines the distance  $\Delta$  of the linear locus from the origin.

Given Eq. (5) and the same point  $\mathbf{x}$  on the given locus, it follows that the inner product  $\mathbf{x}^T \mathbf{u}_{N_e}$  of the given vector  $\mathbf{x}$  with the unit normal eigenaxis  $\mathbf{u}_{N_e}$ :

$$\mathbf{x}^T \mathbf{u}_{N_e} = \Delta,$$

of the given locus, satisfies the distance  $\Delta$  of the linear locus from the origin. Likewise, given Eq. (7), it follows that the normalized, eigen-scaled coordinates of the given point  $\mathbf{x}$

$$\sum_{i=1}^d \cos \alpha_i x_i = \Delta,$$

also satisfy the distance  $\Delta$  of the given locus from the origin.

Finally, given Eq. (3), it follows that the inner product  $\mathbf{x}^T \mathbf{v}$  of the given vector  $\mathbf{x}$  with the normal eigenaxis  $\mathbf{v}$  of the given locus, satisfies the characteristic eigenenergy  $\|\mathbf{v}\|^2$

$$\mathbf{x}^T \mathbf{v} = \|\mathbf{v}\|^2,$$

of the normal eigenaxis  $\mathbf{v}$  of the linear locus.

It is concluded that the uniform properties which are satisfied by all of the points  $\mathbf{x}$  on a linear locus are uniquely determined by the geometric locus of the normal eigenaxis  $\mathbf{v}$  of the linear locus.

Given that the uniform, correlated properties exhibited by the points  $\mathbf{x}$  on a linear locus are uniquely determined by the geometric locus of a normal eigenaxis  $\mathbf{v}$ , it is concluded that all of the points  $\mathbf{x}$  on a line, plane, or hyperplane explicitly and exclusively reference the normal eigenaxis  $\mathbf{v}$  of Eqs (2), (3), and (4).

In summary, it has been demonstrated that the uniform, correlated properties exhibited by any point  $\mathbf{x}$  on any linear locus are functions of the eigen-coordinate locations and the corresponding magnitude and eigenenergy of the normal eigenaxis  $\mathbf{v}$  of the given locus. Thereby, it has been demonstrated that the vector components of a normal eigenaxis provide a characteristic set of eigenaxis locations that effectively specify all forms of linear curves and surfaces. A characteristic set of normal eigenaxis locations will be referred to as eigenloci. It is concluded that a normal eigenaxis  $\mathbf{v}$  coincides as an exclusive and distinctive coordinate axis that effectively characterizes all of the points on a linear locus. Thereby, a normal eigenaxis offers an elegant principal eigen-coordinate system for a linear locus of points.

The next section of the paper will argue that the rich set of geometric properties exhibited by all of the points on a linear locus, which include the normal eigenaxis of a given locus, involve inner product correlations between the *geometric loci of vectors*. Section 5 will examine the idea of the geometric locus of a vector, which will be demonstrated to be the primary building block of regularized, data-driven geometric architectures in Hilbert spaces. It will also be argued that inner product statistics between the geometric loci of training vectors offer a natural *functional glue* for generating learning machine architectures.

## 5 Design of Learning Machine Architectures in Hilbert Spaces

Geometric locus methods restricted to Cartesian coordinate spaces are essentially restricted to static and fixed representations of geometric loci. Indeed, Cartesian coordinate spaces only permit algebraic equations of geometric loci in terms of Euclidean distances between point coordinates and algebraic constraints on point coordinate values. Alternatively, Hilbert spaces permit algebraic systems of geometric loci in terms of correlated algebraic systems of inner product statistics between the geometric loci of vectors, where the magnitude and direction of any given vector determines an endpoint formed by a unique set of point coordinates.

In general, the geometric structures of Euclidean spaces equipped with inner product structures, i.e., Hilbert spaces, are much richer than the geometric structures of Cartesian coordinate spaces Naylor and Sell [1971]. It will be demonstrated that inner product structures in Hilbert spaces actually involve the geometric loci of vectors. Accordingly, the notion of the geometric locus of a point is ill-defined. Given that a geometric locus is a *curve or surface formed by a set of points which possess some uniform property*, it will be shown that the locus of a point actually involves the locus of a vector. To clearly distinguish a vector from a point in the discussion that follows, a vector will be denoted by  $\mathbf{x}$  or  $\mathbf{x} = (x_1, x_2, \dots, x_d)^T$  and a point will be denoted by  $P$  or  $P(x_1, x_2, \dots, x_d)$ .

This section of the paper will demonstrate that a vector  $\tilde{\mathbf{x}} \in \mathbb{R}^d$  is a geometric locus of a directed straight line segment formed by two points  $P_0(0, 0, \dots, 0)$  and  $P_{\tilde{\mathbf{x}}}(\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_d)$ , which are at a distance of

$$\|\tilde{\mathbf{x}}\| = (\tilde{x}_1^2 + \tilde{x}_2^2 + \dots + \tilde{x}_d^2)^{1/2},$$

where each point coordinate  $\tilde{x}_i$  is at a distance of  $\|\tilde{\mathbf{x}}\| \cos \alpha_{ij}$  from the origin  $P_0$ , along the direction of an orthonormal coordinate axis  $\mathbf{e}_j$ , where  $\alpha_{ij}$  is the angle between the vector  $\tilde{\mathbf{x}}$  and an orthonormal coordinate axis  $\mathbf{e}_j$ . This section will also demonstrate how algebraic and geometric structures generated by an inner product statistic between two vectors describe rich topological and algebraic relationships between the geometric loci of two vectors.

### 5.1 Geometric Properties of a Vector

In geometric terms, a vector  $\mathbf{x}$

$$\mathbf{x} = (x_1, x_2, \dots, x_d)^T,$$

is a directed straight line segment that emanates from the point of intersection of coordinate axes  $P_0$

$$P_0(0, 0, \dots, 0),$$

where the values of the coordinates are all zero, commonly known as the origin, such that the endpoint  $P_e$  of the directed straight line segment defines a real specific point

$$P_e(x_1, x_2, \dots, x_d).$$

Thereby, the point coordinates

$$P(x_1, x_2, \dots, x_d),$$

of any given point  $P$  coincide with the components

$$\mathbf{x} = (x_1, x_2, \dots, x_d)^T,$$

of a vector  $\mathbf{x}$ . Given orthogonal coordinate axes, each point coordinate  $x_i$  of  $P$  specifies a scaling for an orthogonal unit coordinate axis  $\mathbf{e}_i$  in  $\mathbb{R}^d$ , where each variable length coordinate axis  $x_i\mathbf{e}_i$  determines a component of a vector  $\mathbf{x} = (x_1, x_2, \dots, x_d)^T$ . For example, Fig. 8 depicts the vectors  $\mathbf{v}$  and  $\mathbf{x}$  in the Euclidean plane  $\mathbb{R}^2$ , where the endpoint of the vector  $\mathbf{v}$  is on the locus of the point  $P_{\mathbf{v}}$ , with point coordinates and vector components  $(x_{v_1}, x_{v_2})$ , and the endpoint of the vector  $\mathbf{x}$  is on the locus of the point  $P_{\mathbf{x}}$ , with point coordinates and vector components  $(x_1, x_2)$ .

So, take any directed straight line segment  $\tilde{\mathbf{x}}$  formed by the points

$$P_{\mathbf{0}}(0, 0, \dots, 0),$$

and

$$P_{\tilde{\mathbf{x}}}(\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_d),$$

where the point  $P_{\tilde{\mathbf{x}}}$  is the endpoint of the vector  $\tilde{\mathbf{x}}$ .

The coordinates of the point  $P_{\tilde{\mathbf{x}}}$  and the components of the vector  $\tilde{\mathbf{x}}$  are both described by the unique, ordered  $d$ -tuple of real numbers:

$$(\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_d),$$

where the Euclidean distance  $D_E$  between  $P_{\mathbf{0}}$  and  $P_{\tilde{\mathbf{x}}}$  is

$$\begin{aligned} D_E(P_{\mathbf{0}}, P_{\tilde{\mathbf{x}}}) &= \left( |0 - \tilde{x}_1|^2 + \dots + |0 - \tilde{x}_d|^2 \right)^{1/2}, \\ &= (\tilde{x}_1^2 + \tilde{x}_2^2 + \dots + \tilde{x}_d^2)^{1/2}. \end{aligned}$$

Now, it makes no sense to speak of the length of a point or to consider the angle between two points. However, any given point  $P_j \in \mathbb{R}^d$  has an enhanced representation as the endpoint of a vector  $\mathbf{x}_j \in \mathbb{R}^d$ . The geometric locus of a vector is defined next.

## 5.2 Geometric Locus of a Vector

Take any given point  $P_{\tilde{\mathbf{x}}}$  which is also the endpoint of a vector  $\tilde{\mathbf{x}}$  in  $\mathbb{R}^d$ . Next, take the standard set of orthonormal basis vectors in  $\mathbb{R}^d$ :

$$\{\mathbf{e}_1 = (1, 0, \dots, 0), \dots, \mathbf{e}_d = (0, 0, \dots, 1)\},$$

and consider the scalar projection of the vector  $\tilde{\mathbf{x}}$  onto the above set of standard basis vectors. The component of the vector  $\tilde{\mathbf{x}}$  along each basis vector  $\mathbf{e}_j$

$$\text{comp}_{\vec{\mathbf{e}}_j}(\vec{\tilde{\mathbf{x}}}) = \|\tilde{\mathbf{x}}\| \cos \alpha_j,$$

where  $\alpha_j$  is the angle between  $\tilde{\mathbf{x}}$  and  $\mathbf{e}_j$ , determines a set of signed magnitudes along the axes of the basis vectors Stewart [2009]

$$(\|\tilde{\mathbf{x}}\| \cos \alpha_1, \|\tilde{\mathbf{x}}\| \cos \alpha_2, \dots, \|\tilde{\mathbf{x}}\| \cos \alpha_d),$$

all of which describe a unique, ordered  $d$ -tuple of geometric loci, where the distance of each point coordinate or vector component  $\tilde{x}_i$  from the origin  $P_o$ , along the axis of the basis vector  $\mathbf{e}_j$ , is  $\|\tilde{\mathbf{x}}\| \cos \alpha_j$ .

## 5.3 Uniform Property Exhibited by Vector Components

Generally speaking, the geometric locus of any vector  $\mathbf{x}_k$  and correlated point  $P_{\mathbf{x}_k}$  is characterized by a unique, ordered  $d$ -tuple of geometric loci:

$$(\|\mathbf{x}_k\| \cos \alpha_{\mathbf{x}_{k1}1}, \|\mathbf{x}_k\| \cos \alpha_{\mathbf{x}_{k2}2}, \dots, \|\mathbf{x}_k\| \cos \alpha_{\mathbf{x}_{kd}d}), \quad (8)$$

where  $(\cos \alpha_{\mathbf{x}_{k1}1}, \dots, \cos \alpha_{\mathbf{x}_{kd}d})$  are the direction cosines of the components  $(x_{k1}, \dots, x_{kd})$  of the vector  $\mathbf{x}_k$  relative to the standard set of orthonormal coordinate axes  $\{\mathbf{e}_j\}_{j=1}^d$ .

Each of the  $d$  point coordinates or vector components  $\{x_{\mathbf{x}_{ki}}\}_{i=1}^d$  are at a distance of  $\|\mathbf{x}_k\| \cos \alpha_{\mathbf{x}_{ki}j}$  from the origin  $P_o$ , along the direction of an orthonormal coordinate axis  $\mathbf{e}_j$ . Thus, each point coordinate or vector component  $x_{\mathbf{x}_{ki}}$  exhibits a characteristic magnitude of  $\|\mathbf{x}_k\| \cos \alpha_{\mathbf{x}_{ki}j}$  along an orthonormal coordinate axis  $\mathbf{e}_j$ .

It is concluded that a vector  $\tilde{\mathbf{x}} \in \mathbb{R}^d$  is a geometric locus of a directed straight line segment formed by two points  $P_o$  and  $P_{\tilde{\mathbf{x}}}$ , which are at a distance of  $\|\tilde{\mathbf{x}}\|$  from each other, where each point coordinate  $\tilde{x}_i$  is at a distance of  $\|\tilde{\mathbf{x}}\| \cos \alpha_{ij}$  from the origin  $P_o$ , along the direction of an orthonormal coordinate axis  $\mathbf{e}_j$ .

It has been demonstrated how a vector provides an enhanced representation of a point. Thereby, it has been demonstrated that the geometric locus of a point is determined by the geometric locus of a vector. It will now be demonstrated how inner product statistics encode a rich set of algebraic and topological relationships between the geometric loci of two vectors.

## 5.4 Inner Product Statistics

The inner product expression  $\mathbf{x}^T \mathbf{x}$  defined by

$$\mathbf{x}^T \mathbf{x} = x_1 x_1 + x_2 x_2 + \cdots + x_d x_d,$$

generates the norm  $\|\mathbf{x}\|$  of the vector  $\mathbf{x}$

$$\|\mathbf{x}\| = (x_1^2 + x_2^2 + \cdots + x_d^2)^{1/2},$$

which determines the Euclidean distance between the endpoint of  $\mathbf{x}$  and the origin, whereby the norm  $\|\mathbf{x}\|$  measures the length of the vector  $\mathbf{x}$ , which indicates the magnitude of  $\mathbf{x}$ . The Euclidean space  $\mathbb{R}^d$  equipped with a norm  $\|\mathbf{x}\|$  that is generated by an inner product  $\mathbf{x}^T \mathbf{x}$  is a Hilbert space Naylor and Sell [1971].

The inner product function  $\mathbf{x}^T \mathbf{y}$  also determines the angle between two vectors  $\mathbf{x}$  and  $\mathbf{y}$  in  $\mathbb{R}^d$ . Given any two vectors  $\mathbf{x}$  and  $\mathbf{y}$ , the inner product expression

$$\mathbf{x}^T \mathbf{y} = x_1 y_1 + x_2 y_2 + \cdots + x_d y_d, \quad (9)$$

is also given by the expression

$$\mathbf{x}^T \mathbf{y} = \|\mathbf{x}\| \|\mathbf{y}\| \cos \theta, \quad (10)$$

where  $\theta$  is the angle between the vectors  $\mathbf{x}$  and  $\mathbf{y}$ . If  $\theta = 90^\circ$ ,  $\mathbf{x}$  and  $\mathbf{y}$  are said to be perpendicular to each other. Accordingly, the inner product expression in Eq. (10) allows us to describe vectors which are orthogonal or perpendicular to each other Naylor and Sell [1971]. Two vectors  $\mathbf{x}$  and  $\mathbf{y}$  are said to be orthogonal to each other if

$$\mathbf{x}^T \mathbf{y} = 0,$$

which is denoted by  $\mathbf{x} \perp \mathbf{y}$ .

It will now be demonstrated how the algebraic relationships in Eqs (9) and (10) are derived from second-order distance statistics between the geometric loci of two vectors. Second-order distance statistics will be shown to encode rich algebraic and topological relationships between the geometric loci of vectors.

## 5.5 Second-order Distance Statistics Between Loci of Vectors

The algebraic relationship

$$v^T \nu = \|v\| \|\nu\| \cos \varphi,$$

between two vectors  $v$  and  $\nu$  in Hilbert space can be derived by using the law of cosines Lay [2006]:

$$\|v - \nu\|^2 = \|v\|^2 + \|\nu\|^2 - 2 \|v\| \|\nu\| \cos \varphi, \quad (11)$$

which reduces to

$$\begin{aligned}\|v\| \|\nu\| \cos \varphi &= v_1 \nu_1 + v_2 \nu_2 + \dots + v_d \nu_d, \\ &= v^T \nu, \\ &= \nu^T v.\end{aligned}$$

This indicates that the inner product statistic  $v^T \nu$  or  $\|v\| \|\nu\| \cos \varphi$  determines the length  $\|v - \nu\|$  of the vector from  $\nu$  to  $v$ , which is the distance between the endpoints of  $v$  and  $\nu$ .

It follows that the inner product statistic between any two vectors  $v$  and  $\nu$  in Hilbert space

$$\begin{aligned}v^T \nu &= v_1 \nu_1 + v_2 \nu_2 + \dots + v_d \nu_d, \\ &= \|v\| \|\nu\| \cos \varphi,\end{aligned}$$

determines the distance between the geometric loci of  $v$  and  $\nu$ .

It is concluded that the algebraic relationships defined within Eq. (11) describe a rich set of topological relationships between the geometric loci of two vectors. Figure 9 depicts the rich set of correlated algebraic and topological structures encoded within an inner product statistic of the geometric loci of two vectors.

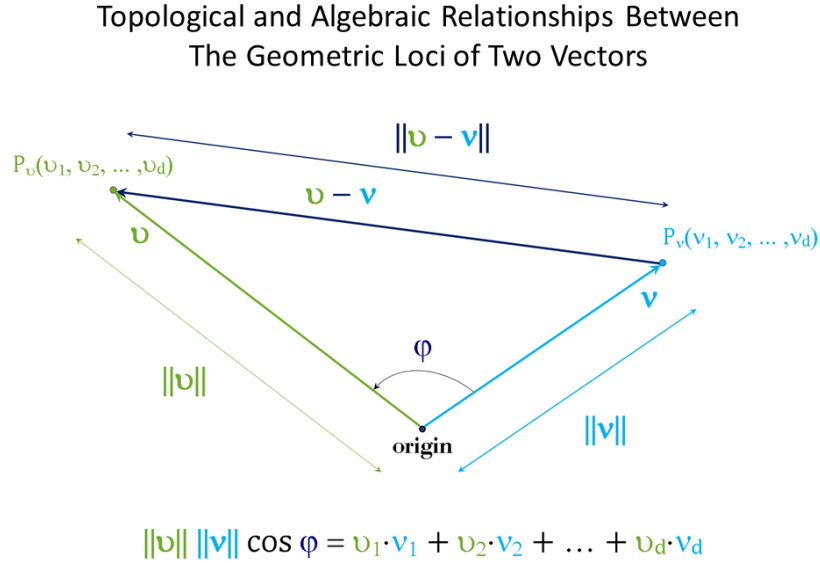


Figure 9: Illustration of the rich set of algebraic and topological relationships encoded within the inner product statistic  $v^T \nu$  of the geometric loci of the vectors  $v$  and  $\nu$ .



Equation (11) also determines the component of a vector along another vector, which is also known as a scalar projection. The algebraic and geometric nature of scalar projections are examined next.

## 5.6 Signed Magnitudes of Vector Projections

The inner product statistic

$$\mathbf{x}^T \mathbf{y} = \|\mathbf{x}\| \|\mathbf{y}\| \cos \theta,$$

can be interpreted as the length  $\|\mathbf{x}\|$  of  $\mathbf{x}$  times the scalar projection of  $\mathbf{y}$  onto  $\mathbf{x}$

$$\mathbf{x}^T \mathbf{y} = \|\mathbf{x}\| \times [\|\mathbf{y}\| \cos \theta], \quad (12)$$

where the scalar projection of  $\mathbf{y}$  onto  $\mathbf{x}$ , also known as the component of  $\mathbf{y}$  along  $\mathbf{x}$ , is defined to be the signed magnitude of the vector projection

$$\|\mathbf{y}\| \cos \theta,$$

where  $\theta$  is the angle between  $\mathbf{x}$  and  $\mathbf{y}$  Stewart [2009]. Scalar projections are denoted by  $\text{comp}_{\vec{\mathbf{x}}}(\vec{\mathbf{y}})$ , where  $\text{comp}_{\vec{\mathbf{x}}}(\vec{\mathbf{y}}) < 0$  if  $\pi/2 < \theta \leq \pi$ .

The scalar projection statistic also satisfies the inner product relationship

$$\begin{aligned} \|\mathbf{y}\| \cos \theta &= \frac{\mathbf{x}^T \mathbf{y}}{\|\mathbf{x}\|} \\ &= \left( \frac{\mathbf{x}}{\|\mathbf{x}\|} \right)^T \mathbf{y}, \end{aligned}$$

between the unit vector  $\frac{\mathbf{x}}{\|\mathbf{x}\|}$  and  $\mathbf{y}$ .

Figure 10 depicts the geometric nature of scalar projections for acute and obtuse angles between vectors. Scalar projection statistics determine signed magnitudes along the axes of given vectors.

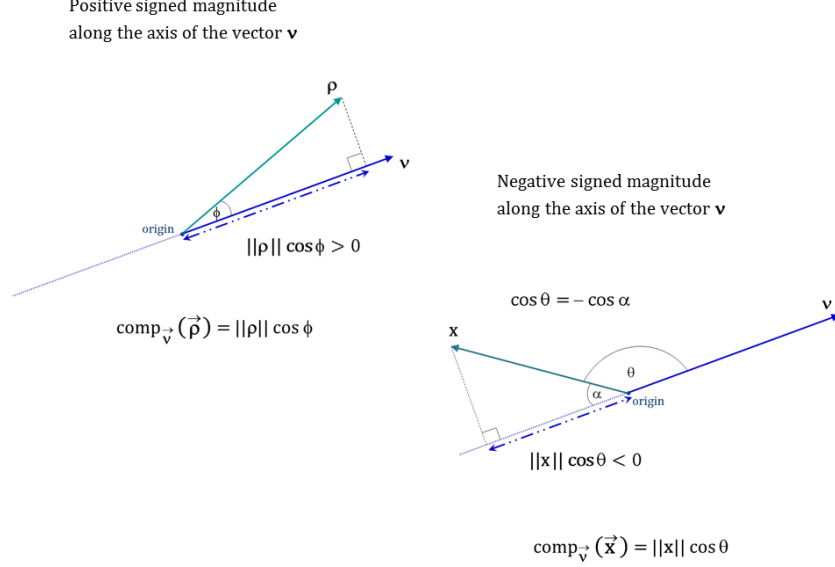


Figure 10: Illustration of how scalar projection statistics determine components (signed magnitudes) of vectors along the axis of a given vector.

The findings presented in this paper will demonstrate how the algebraic and topological relationships encoded within inner product statistics offer a natural *functional glue* for learning machine architectures.

## 5.7 Designing Functional Glue for Learning Machine Architectures

This paper will demonstrate how algebraic systems of properly specified, correlated inner product statistics between training data give rise to regularized, data-driven geometric architectures which encode robust decision statistics for complex discrimination tasks. The remaining sections of this paper will develop a regularized, data-driven geometric architecture, which describes linear decision boundaries for overlapping and non-overlapping data distributions, that is determined by correlated sets of dual, i.e., primal and dual, principal (normal) eigenaxis components, all of which are jointly and symmetrically located in primal and dual, correlated Hilbert spaces. It will be shown that data-driven sets of primal and dual normal eigenaxis components encode robust likelihood ratios for complex discrimination tasks. It will also be demonstrated that eigenenergies of data-driven sets of primal and dual normal eigenaxis components satisfy the law of cosines in a surprisingly elegant and symmetric manner.

The analyses that follow will make extensive use of inner product and scalar projection statistics. Inner product and scalar projection statistics will be shown

to provide a natural functional glue for adaptable geometric architectures. Regularized, adaptable geometric architectures which encode relevant aspects of statistical decision systems will be demonstrated to be the principal foundation of learning machine architectures.

It has previously been argued that Cartesian coordinate spaces only permit static and fixed descriptions of a geometric locus of points. The analyses that follow will demonstrate how algebraic systems of correlated inner product statistics between training vectors in dual, correlated Hilbert spaces generate robust, data-driven, symmetrical geometric architectures that represent statistical decision systems. Thereby, it will be demonstrated how collections of training data are transformed into regularized geometric architectures which encode relevant geometric and statistical aspects of statistical decision systems.

Naylor and Sell note that a truly amazing number of problems in engineering and science can be fruitfully treated with geometric methods in Hilbert space Naylor and Sell [1971]. *This paper will use geometric and statistical methods in dual, correlated Hilbert spaces to solve the long-standing problem of learning robust or optimal linear decision boundaries for overlapping sets of data.*

The set of analyses which follow involve the examination of dual, interconnected, symmetrical geometric architectures which are generated by algebraic systems of correlated inner product statistics between training vectors in dual, correlated Hilbert spaces in  $\mathbb{R}^d$  and  $\mathbb{R}^N$ . The analyses will demonstrate how symmetrical, interconnected geometric architectures in dual, correlated Hilbert spaces provide the geometric basis of the statistical representation of a primal normal eigenaxis in a Wolfe dual eigenspace. The analyses will use all of the algebraic and geometric properties of the normal eigen-coordinate system outlined earlier, to examine how robust estimates of constrained normal eigen-coordinate locations provide robust, stable, and optimal statistical representations of linear decision boundaries. The analyses will demonstrate that robust estimates of constrained normal eigenaxis components provide optimal statistical descriptions of linear decision boundaries for normally distributed training data with common covariance matrices. The analyses will also demonstrate that robust estimates of constrained normal eigen-coordinate locations provide robust or optimal estimates of linear decision boundaries for training data drawn from various distributions, including completely overlapping data distributions.

More generally, it can be demonstrated that the eigen-coordinate locations of the principal eigenaxis of any given second-order curve or surface offer a characteristic set of eigenloci that specify the given curve or surface. An upcoming paper will consider how principal eigenaxes provide exclusive, intrinsic coordinate axes for the geometric loci of  $d$ -dimensional circles, ellipses, hyperbolae, and parabolas. The paper will examine how robust statistical representations of constrained principal eigen-coordinate locations provide the primary statistical basis for second-order decision boundary estimates. Indeed, robust estimates of constrained principal eigen-coordinate locations can be shown to describe optimal binary decision boundaries for all forms of normally distributed training data.

A high level description of the linear SVM method in a geometric locus

framework is outlined next. The outline is intended to motivate the development of a computer-implemented method that effectively hardwires the geometric locus of a normal eigenaxis into linear kernel SVM architectures. The term eigenlocus is used to refer to the locus of a normal eigenaxis.

## 5.8 Hardwiring the Eigenlocus of a Normal Eigenaxis into Linear Kernel SVM Architectures

So far, the principal eigenaxis of linear curves and surfaces has been identified and given the name *normal eigenaxis*. It has been established that normal eigenaxes of linear curves and surfaces are major intrinsic axes that coincide as exclusive reference axes. It has been demonstrated that all of the points on a linear locus are specified by the eigen-coordinate locations and corresponding magnitude and eigenenergy of its normal eigenaxis. Therefore, given the normal eigenaxis of any linear locus of points, it follows that the geometric locus of any given normal eigenaxis has a distinctive geometric configuration that is specified by a characteristic set of eigen-coordinates, all of which jointly determine the characteristic location and eigenenergy of the normal eigenaxis. Because normal eigenaxes of distinct linear curves or surfaces possess invariant and distinctive geometric locations, it follows that the location of a normal eigenaxis is an invariant and hardwired geometric property of lines, planes, and hyperplanes. Thus, *the important generalizations for lines, planes, and hyperplanes, are hardwired into (encoded within) the geometric locus of a normal eigenaxis*. Furthermore, the normal eigenaxis of any given linear locus satisfies the linear locus in terms of its eigenenergy. Thereby, *the fundamental property of a normal eigenaxis is its eigenenergy*.

Clearly, then, the primary curve of interest for learning linear decision boundaries is a normal eigenaxis. This implies that the important generalizations for linear decision boundaries involve an estimation process that encodes the geometric locus of a normal eigenaxis within learning machine architectures. Accordingly, robust and optimal estimates of linear decision boundaries must be based on effective statistical representations of constrained normal eigen-coordinate locations of unknown linear decision boundaries. The remaining portions of this paper will refer to a normal eigenaxis as a normal eigenlocus. The paper will use the term statistical eigenlocus to refer to a statistical estimate of a normal eigenlocus. The paper will use the term strong dual normal eigenlocus to refer to joint, statistical eigenlocus estimates in dual, correlated Hilbert spaces. Given all of the above assumptions, the remaining sections of this paper will develop algebraic and statistical expressions for a strong dual normal eigenlocus of normal eigenaxis components that provides a robust statistical representation of the constrained normal eigen-coordinate locations of an unknown linear decision boundary.

### Definition

A strong dual normal eigenlocus of normal eigenaxis components is a normal eigenlocus of a linear decision boundary which encodes the constrained eigen-coordinate locations of an unknown normal eigenaxis. A strong dual normal eigenlocus satisfies a linear decision boundary in terms of a critical minimum, i.e., a total allowed, eigenenergy. The term eigenlocus will be used to refer to the location of a normal eigenaxis component on a normal eigenlocus, or to the location of a normal eigenlocus; the context will make the meaning clear.

## 5.9 A Strong Dual Normal Eigenlocus of Normal Eigenaxis Components

Consider a normal eigenlocus of a linear decision boundary formed by a strong dual normal eigenlocus of normal eigenaxis components, all of which are eigen-scaled extreme data points of large covariance, all of which encode a robust likelihood ratio, each of which determines an eigen-transformed principal location of large covariance. A strong dual normal eigenlocus of a separating line, plane, or hyperplane will be shown to satisfy a linear decision boundary in terms of a critical minimum eigenenergy. The analyses that follow will use Eqs (2), (3), (4), and (7), along with the correlated algebraic and geometric properties of these equations, to demonstrate how the constrained normal eigen-coordinate locations of an unknown linear decision boundary can be estimated from training data by means of a properly specified strong dual normal eigenlocus of normal eigenaxis components, each of which encodes the probability of finding an extreme data point in a particular region of  $\mathbb{R}^d$ . The analyses will demonstrate that the eigenlocus of each normal eigenaxis component encodes an eigen-balanced first and second order statistical moment about the locus of an extreme training point, which is shown to determine the likelihood of finding the extreme data point in a particular region of  $\mathbb{R}^d$ . First and second order statistical moments, which involve unidirectional estimates of joint variations between a given vector and a collection of training data, provide an estimate of how the components of the given vector are distributed within the training data. The eigen-balanced first and second order statistical moments encoded within a strong dual normal eigenlocus of normal eigenaxis components will be shown to describe a large number of data distributions. A strong dual normal eigenlocus of normal eigenaxis components will be formally referred to as a strong dual normal eigenlocus.

## 5.10 High Level Overview of a Strong Dual Normal Eigenlocus

Let the term *strong dual normal eigenlocus* refer to a dual statistical eigenlocus of normal eigenaxis components which delineates and satisfies three, symmetrical linear partitioning curves or surfaces. A strong dual normal eigenlocus satisfies three, symmetrical linear partitioning curves or surfaces in terms of a

critical minimum eigenenergy. The normal eigenaxis components on a strong dual normal eigenlocus are analogous to the sinusoidal components of a Fourier series. All of the sinusoidal components of a given Fourier series have such amplitudes and phases that they sum to an approximation of a distinct periodic function or signal Lathi [1998]. Likewise, all of the normal eigenaxis components on a strong dual normal eigenlocus have such magnitudes and directions that they sum to an estimate of a normal eigenaxis of three, characteristic and symmetrical linear partitioning curves or surfaces. How such a statistical balancing feat can be routinely accomplished is a central idea of this paper. This paper will examine how achieving this type of statistical equilibrium involves identifying and exploiting effective statistical representations for constrained eigen-coordinate locations of unknown normal eigenaxes of unknown linear decision boundaries.

The sections that follow will examine an estimation process that transforms two sets of pattern vectors, generated by any two probability distributions whose expected values and covariance structures do not vary over time, into a dual statistical eigenlocus of normal eigenaxis components, all of which are jointly and symmetrically located in dual and primal, correlated Hilbert spaces, all of which encode a robust likelihood ratio, all of which jointly describe symmetrical, linear subspaces of  $\mathbb{R}^N$  and  $\mathbb{R}^d$ , each of which determines an eigen-balanced, pointwise covariance estimate of an extreme data point located between two data distributions in  $\mathbb{R}^d$ . All of the normal eigenaxis components on a strong dual normal eigenlocus jointly determine a statistical decision system, of three, symmetrical linear partitioning curves or surfaces in  $\mathbb{R}^d$ , that delineates bipartite, congruent geometric regions of large covariance located between two data distributions in  $\mathbb{R}^d$ , such that the congruent geometric regions of large covariance delineate regions of data distribution overlap for overlapping distributions. The resultant loci of points on all three linear curves or surfaces in  $\mathbb{R}^d$  exclusively reference the dual statistical eigenlocus of normal eigenaxis components. Likelihoods encoded within the eigenloci of all of the normal eigenaxis components specify the stochastic behavior of a statistical decision system.

The next section will begin to develop the primal and the Wolfe dual normal eigenlocus equations of a probabilistic, binary linear classification system.

## 6 The Primal and the Wolfe Dual Normal Eigenlocus Equations of a Probabilistic Binary Linear Classification System

The eigenlocus equations of a strong dual normal eigenlocus are commonly referred to as soft margin linear support vector machines. The analyses that follow will show that the subset of weighted training points, commonly called support vectors, form a dual statistical eigenlocus of eigen-transformed extreme training points, all of which jointly determine a statistical decision system of three, symmetrical linear partitioning curves or surfaces in  $\mathbb{R}^d$ . It will be demon-

strated that each support vector is an eigen-scaled extreme training point that determines a well-proportioned (eigen-balanced) and properly-positioned normal eigenaxis component on a strong dual normal eigenlocus. To wit, it will be demonstrated that support vectors are major eigenaxis components of an intrinsic reference axis of a linear decision boundary and bilaterally symmetrical borders.

Finding a separating line, plane, or hyperplane requires estimating the normal eigenlocus of a linear decision boundary and the bilaterally symmetrical borders which bound it. The analyses that follow will define the complete statistical system of a strong dual normal eigenlocus, and thereby will identify a probabilistic linear discriminant function that is Bayes' optimal for common covariance data.

The study begins with the eigenlocus equation of a primal normal eigenlocus.

## 6.1 Eigenlocus Equation of a Primal Normal Eigenlocus

The strong dual normal eigenlocus  $\tau$  of a separating line, plane, or hyperplane is estimated by solving an inequality constrained optimization problem:

$$\begin{aligned} \min \Psi(\tau) &= \|\tau\|^2/2 + C/2 \sum_{i=1}^N \xi_i^2 \\ \text{s.t. } y_i(\mathbf{x}_i^T \tau + \tau_0) &\geq 1 - \xi_i, \quad \xi_i \geq 0, \quad i = 1, \dots, N, \end{aligned} \quad (13)$$

where  $\tau$  is a  $d \times 1$  constrained primal normal eigenlocus of three, symmetrical linear partitioning curves or surfaces,  $C$  and  $\xi_i$  are regularization parameters,  $y_i$  are training set labels (if  $\mathbf{x}_i \in H_1$ , assign  $y_i = 1$ ; otherwise, assign  $y_i = -1$ ), and  $\tau_0$  is a function of  $\tau$ , extreme training points  $\mathbf{x}_{i*}$ , and training set labels  $y_i$ .

It will be demonstrated that Eq. (13) provides the primal (elemental) specification of a linear decision boundary that is bounded by bilaterally symmetrical decision borders. Any given linear decision boundary is centrally and symmetrically positioned between any two given data distributions, such that the linear decision borders span symmetrical regions of large covariance. The strong dual solution of Eq. (13) involves solving a complementary and essential optimization problem that determines the fundamental unknowns of Eq. (13). It is claimed that the actual unknowns in Eq. (13) are the constrained eigen-coordinate locations of a normal eigenaxis  $\mathbf{v}$  that delineates and satisfies three, symmetrical lines, planes, or hyperplanes, all of which jointly delineate a symmetrical partitioning of a feature space in  $\mathbb{R}^d$ .

It will be shown that the locations of the normal eigenaxis components on  $\tau$  provide estimates for the constrained eigen-coordinate locations of  $\mathbf{v}$ . It will also be demonstrated that  $\tau$  provides an exclusive, intrinsic reference axis for any given linear decision boundary and decision borders.

Moreover, it will be shown that the strong dual solution of Eq. (13) determines a statistical equilibrium point, i.e., the eigenlocus of  $\tau$ , such that a constrained discriminant function  $\tau^T \mathbf{x} + \tau_0$  delineates a centrally and symmetrically positioned, bipartite geometric region of constrained, constant, and equal

widths, that spans a region of high variability (large covariance) between two data distributions, whereby the bipartite, congruent regions of large covariance delineate symmetrical regions of data distribution overlap for overlapping data distributions.

It will also be shown that a strong dual normal eigenlocus  $\tau$  satisfies three, symmetrical linear partitioning curves or surfaces in terms of a critical minimum eigenenergy. Thereby, it will be shown that a strong dual normal eigenlocus  $\tau$  possesses a critical minimum eigenenergy which is the fundamental geometric and statistical property of  $\tau$ .

## 6.2 The Critical Minimum Eigenenergy Constraint on $\tau$

Given Eq. (13) and the assumptions outlined above, it follows that  $N$  primal normal eigenlocus equations must be satisfied:

$$y_i (\mathbf{x}_i^T \tau + \tau_0) \geq 1 - \xi_i, \quad \xi_i \geq 0, \quad i = 1, \dots, N,$$

such that a constrained primal normal eigenlocus  $\tau$  satisfies a critical minimum eigenenergy constraint:

$$\gamma(\tau) = \|\tau\|_{\min_c}^2, \quad (14)$$

where the total allowed eigenenergy  $\|\tau\|_{\min_c}^2$  of  $\tau$  is the fundamental geometric and statistical property of  $\tau$ . It will be shown that  $\tau$  possesses a critical minimum eigenenergy

$$\begin{aligned} \|\tau\|_{\min_c}^2 &= \|\tau_1 - \tau_2\|_{\min_c}^2, \\ &= \|\tau_1\|_{\min_c}^2 + \|\tau_2\|_{\min_c}^2 - 2 \|\tau_1\| \|\tau_2\| \cos \theta_{\tau_1 \tau_2}, \end{aligned}$$

where  $\tau_1$  and  $\tau_2$  are components of  $\tau$

$$\tau = \tau_1 - \tau_2,$$

such that the total allowed eigenenergies of  $\tau_1$  and  $\tau_2$  are effectively balanced by means of a symmetric equalizer statistic  $\nabla_{eq}$

$$\left( \|\tau_1\|_{\min_c}^2 - \|\tau_1\| \|\tau_2\| \cos \theta_{\tau_1 \tau_2} \right) + \nabla_{eq} \Leftrightarrow \left( \|\tau_2\|_{\min_c}^2 - \|\tau_1\| \|\tau_2\| \cos \theta_{\tau_1 \tau_2} \right) - \nabla_{eq},$$

in relation to a centrally located statistical fulcrum  $f_s$ . It will also be demonstrated that the critical minimum eigenenergy  $\|\tau\|_{\min_c}^2$  exhibited by  $\tau$  determines a statistical equilibrium point which encodes a robust likelihood ratio for all data distributions.

Figure 11 illustrates the algebraic, geometric, and statistical nature of the remarkable statistical balancing feat that is routinely accomplished by solving the inequality constrained optimization problem in Eq. (13).



## Statistical Equilibrium Point of $\tau$

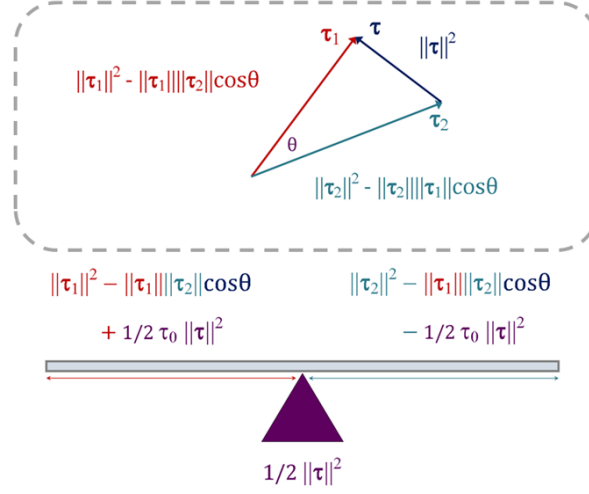


Figure 11: Illustration of the algebraic and topological constraints which determine a strong dual normal eigenlocus equilibrium point. The center of eigenenergy of  $\tau$  is subjected to equal and opposite eigenenergies, so that a strong dual normal eigenlocus  $\tau = \tau_1 - \tau_2$  achieves a state of statistical equilibrium.

### 6.3 The Strong Dual Normal Eigenlocus Equilibrium Point

Section 17 will show that the critical minimum eigenenergy constraint on  $\tau$  determines a strong dual normal eigenlocus equilibrium point, i.e., the location or eigenlocus of  $\tau$ , whereby a constrained discriminant function  $\tau^T \mathbf{x} + \tau_0$  delineates the positions of three, symmetrical linear partitioning curves or surfaces. Section 18 will demonstrate that the total allowed eigenenergy and statistical equilibrium point of  $\tau$  is specified by likelihood statistics encoded within correlated normal eigenaxis components on a Wolfe dual normal eigenlocus. Denote a Wolfe dual normal eigenlocus by  $\psi$  and its eigenlocus equation by  $\max \Xi(\psi)$ .

Let  $\psi$  be a Wolfe dual of  $\tau$ , such that proper and effective strong duality relationships exist between the algebraic systems of  $\min \Psi(\tau)$  and  $\max \Xi(\psi)$ . Thereby, let  $\psi$  be related with  $\tau$  in a symmetrical manner that specifies the locations of the normal eigenaxis components on  $\tau$ . The Wolfe dual normal eigenlocus  $\psi$  is important for the following reasons.

### 6.4 Why the Wolfe Dual Normal Eigenlocus Matters

Duality relationships for Lagrange multiplier problems are based on the premise that it is the Lagrange multipliers which are the fundamental unknowns asso-

ciated with a constrained problem. Dual methods solve an alternate problem, termed the dual problem, whose unknowns are the Lagrange multipliers of the first problem, termed the primal problem. Once the Lagrange multipliers are known, the solution to a primal problem can be determined Luenberger [2003].

It is assumed that the real unknowns associated with the inequality constrained optimization problem in Eq. (13) are the constrained eigen-coordinate locations of a normal eigenaxis  $\mathbf{v}$  that delineates the geometric configuration of a linear decision boundary and the widths of its decision borders. It is also assumed that a normal eigenaxis  $\mathbf{v}$  satisfies a linear decision boundary and its decision borders in terms of a critical minimum, i.e., a total allowed, eigenenergy. The main issue concerns how the constrained eigen-coordinate locations of a normal eigenaxis  $\mathbf{v}$  are determined.

It will be demonstrated that the constrained eigen-coordinate locations of  $\mathbf{v}$  are estimated by the locations of normal eigenaxis components on a constrained primal normal eigenlocus  $\tau$ , all of which are effectively determined by the locations of normal eigenaxis components on a Wolfe dual normal eigenlocus  $\psi$ . To wit, it will be shown that the constrained eigen-coordinate locations of  $\mathbf{v}$  are essentially determined by the eigenloci of normal eigenaxis components on a Wolfe dual normal eigenlocus  $\psi$ . It will be demonstrated that the eigenloci of the Wolfe dual normal eigenaxis components on  $\psi$  determine critical minimum eigenenergies for  $\psi$  and  $\tau$ , both of which jointly determine a statistical equilibrium point for  $\tau$ . Section 17 will define a strong dual normal eigenlocus equilibrium point for which the critical minimum eigenenergies of  $\tau = \tau_1 - \tau_2$  satisfy a linear decision boundary and its decision borders.

## 6.5 Fundamental Unknowns for Strong Dual Normal Eigenlocus Estimates

For the problem of strong dual normal eigenlocus estimates, the Lagrange multipliers method introduces a Wolfe dual normal eigenlocus  $\psi$  of normal eigenaxis components, for which the Lagrange multipliers  $\{\psi_i\}_{i=1}^N$  are the magnitudes or lengths of a set of Wolfe dual normal eigenaxis components  $\{\psi_i \vec{e}_i\}_{i=1}^N$ , and finds extrema for the restriction of a primal normal eigenlocus  $\tau$  to a Wolfe dual eigenspace. Accordingly, the fundamental unknowns associated with Eq. (13) are the magnitudes or lengths of the Wolfe dual normal eigenaxis components on  $\psi$ . It will be shown that each Lagrange multiplier provides an eigen-scale that determines the length of a correlated, constrained primal normal eigenaxis component on  $\tau$ .

Because Eq. (13) is a convex programming problem, the theorem for convex duality guarantees some type of equivalence and corresponding symmetry between a constrained primal normal eigenlocus and its Wolfe dual. Strong duality holds between the algebraic systems of  $\min \Psi(\tau)$  and  $\max \Xi(\psi)$ , so that the duality gap between the constrained primal and the Wolfe dual normal eigenlocus solution is zero Luenberger [1969], Nash and Sofer [1996], Fletcher [2000], Luenberger [2003].

This paper will demonstrate how strong duality relationships between the algebraic systems of  $\min \Psi(\tau)$  and  $\max \Xi(\psi)$  ensure that the total allowed eigenenergy  $\|\psi\|_{\min_c}^2$  and the eigenlocus of  $\psi$  are symmetrically related to the total allowed eigenenergy  $\|\tau\|_{\min_c}^2$  and the statistical equilibrium point of  $\tau$ . These relationships will be defined in Sections 12, 14, 15, 16, and 17. Section 18 will show that the total allowed eigenenergy  $\|\tau\|_{\min_c}^2$  of  $\tau$  describes the likelihood of finding data points in particular regions of  $\mathbb{R}^d$ . Section 18 will also demonstrate that the statistical equilibrium point of  $\tau$  encodes Bayes' likelihood ratio for common covariance data and a robust likelihood ratio for all other data.

This paper will also demonstrate how strong duality relationships between the algebraic systems of  $\min \Psi(\tau)$  and  $\max \Xi(\psi)$  ensure that the geometric architecture described by  $\max \Xi(\psi)$  is symmetrically related to the geometric architecture of the statistical decision system described by  $\min \Psi(\tau)$ . The strong duality relationships between the algebraic systems of  $\min \Psi(\tau)$  and  $\max \Xi(\psi)$  will be shown to determine symmetrical linear partitioning systems in  $\mathbb{R}^N$  and  $\mathbb{R}^d$ , which jointly determine a learning machine architecture in  $\mathbb{R}^d$  that exhibits a surprising amount of bilateral symmetry for arbitrary data distributions.

The term *strong dual* will frequently be used to emphasize the joint geometric and statistical properties exhibited by a constrained primal and a Wolfe dual normal eigenlocus. The matrix version of the Wolfe dual normal eigenlocus equation is summarized below.

## 6.6 The Wolfe Dual Normal Eigenlocus of a Separating Hyperplane

The complementary and essential normal eigenlocus estimate, which is specified by the Wolfe dual normal eigenlocus of Eqs (39) or (40), involves finding normal eigenaxis components that are determined by the minimization of a constrained quadratic form

$$\max \Xi(\psi) = \mathbf{1}^T \psi - \frac{\psi^T \mathbf{Q} \psi}{2},$$

subject to the constraints  $\psi^T \mathbf{y} = 0$  and  $\psi_i \geq 0$ , where  $\mathbf{Q} \triangleq \epsilon \mathbf{I} + \tilde{\mathbf{X}} \tilde{\mathbf{X}}^T$ , the matrix  $\tilde{\mathbf{X}} \triangleq \mathbf{D}_y \mathbf{X}$ ,  $\mathbf{D}_y$  is an  $N \times N$  diagonal matrix of training labels  $y_i$  and the  $N \times d$  data matrix is  $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N)^T$ . The eigenlocus equation of  $\max \Xi(\psi)$  will be derived in sections that follow.

The analyses that follow will examine how the strong duality relationships between the algebraic systems of  $\min \Psi(\tau)$  and  $\max \Xi(\psi)$  determine a strong dual normal eigenlocus of symmetrical linear partitioning systems in  $\mathbb{R}^N$  and  $\mathbb{R}^d$ . The critical minimum eigenenergy  $\|\psi\|_{\min_c}^2$  of  $\psi$  will be shown to be symmetrically related to the critical minimum eigenenergy  $\|\tau\|_{\min_c}^2$  of  $\tau$ . Thereby, this paper will demonstrate how the geometric configuration of a Wolfe dual normal eigenlocus  $\psi$  determines the geometric configuration of a constrained primal normal eigenlocus  $\tau$ .

## 6.7 Symmetrical Linear Partitioning Systems in $\mathbb{R}^N$ and $\mathbb{R}^d$

Equation (13) and the existence of strong duality relationships between the algebraic systems of  $\min \Psi(\tau)$  and  $\max \Xi(\psi)$  indicate that three, symmetrical hyperplane partitioning surfaces are delineated by the constrained quadratic form denoted by  $\max \Xi(\psi)$ . Given these assumptions and Eqs (4) and (7), it follows that any point on a hyperplane surface possesses a set of normalized, eigen-scaled coordinates which satisfy the distance of the hyperplane surface from the origin, where each distance is determined by a correlated constraint on the constrained discriminant function of Eq. (22). Section 11 will show that the geometric configurations of all three hyperplane surfaces are an inherent function of the inner product elements of the Gram matrix  $\mathbf{Q}$  associated with the constrained quadratic form in the equation denoted by  $\max \Xi(\psi)$  or Eq. (40). Sections 12 - 16 will examine how a Wolfe dual normal eigenlocus  $\psi$  delineates and satisfies three, symmetrical hyperplane partitioning surfaces in terms of a critical minimum eigenenergy constraint.

Sections 14 and 15 will examine how the Lagrange multipliers of a primal normal eigenlocus problem provide an estimate of constrained normal eigen-coordinate locations that implicitly delineate a separating hyperplane in  $\mathbb{R}^N$  which is effectively bounded by bilateral symmetrical hyperplane borders. Sections 14 and 15 will show how each of the normal eigenaxis components on  $\psi \in \mathbb{R}^N$  encodes an eigen-scale that determines a critical length for a symmetrical normal eigenaxis component on  $\tau \in \mathbb{R}^d$ , such that  $\tau$  delineates a statistical decision system of three, symmetrical linear partitioning curves or surfaces in  $\mathbb{R}^d$ .

Figure 12 depicts a high level overview of the symmetrical relationships between a constrained primal normal eigenlocus  $\tau$  and its Wolfe dual  $\psi$ , where symmetry involves regularized correlations between the critical minimum eigenenergies of  $\tau$  and  $\psi$ , which jointly determine the statistical equilibrium point  $\tau$  that is satisfied by  $\tau$  and  $\psi$ , all of which jointly determine regularized correlations between dual, linear partitioning systems in  $\mathbb{R}^d$  and  $\mathbb{R}^N$ .

Denote the set of hyperplane partitioning surfaces in  $\mathbb{R}^N$  by  $H_0$ ,  $H_{+1}$ , and  $H_{-1}$ , where the critical minimum eigenenergy  $\|\psi\|_{\min_c}^2$  of  $\psi$  satisfies  $H_0$ ,  $H_{+1}$ , and  $H_{-1}$ . Let the set of linear partitioning surfaces in  $\mathbb{R}^d$ , which are determined by constraining the expression  $\tau^T \mathbf{x} + \tau_0$  to be equal to 0, +1, and -1, be denoted by  $D_0(\mathbf{x})$ ,  $D_{+1}(\mathbf{x})$ , and  $D_{-1}(\mathbf{x})$ , where the critical minimum eigenenergy  $\|\tau\|_{\min_c}^2$  of  $\tau$  satisfies  $D_0(\mathbf{x})$ ,  $D_{+1}(\mathbf{x})$ , and  $D_{-1}(\mathbf{x})$ . Figure 12 illustrates how strong duality relationships between the algebraic systems of  $\min \Psi(\tau)$  and  $\max \Xi(\psi)$  ensure that the geometric configurations of the hyperplane partitioning surfaces  $H_0$ ,  $H_{+1}$ , and  $H_{-1}$  in  $\mathbb{R}^N$  regulate the geometric configurations of the linear partitioning surfaces  $D_0(\mathbf{x})$ ,  $D_{+1}(\mathbf{x})$ , and  $D_{-1}(\mathbf{x})$  in  $\mathbb{R}^d$ . Accordingly, the geometric configuration of a separating hyperplane  $H_0$  is symmetrically related to the geometric configuration of a linear decision boundary  $D_0(\mathbf{x})$ . Likewise, the geometric configurations of the hyperplane decision borders  $H_{+1}$  and  $H_{-1}$  are symmetrically related to the geometric configurations of the linear

decision borders  $D_{+1}(\mathbf{x})$ , and  $D_{-1}(\mathbf{x})$ .

### Overview of the Symmetrical Relationships Between a Constrained Primal Normal Eigenlocus and its Wolfe Dual

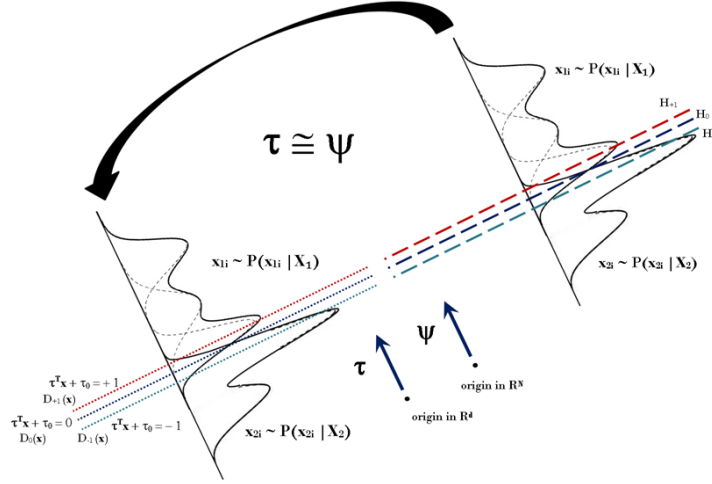


Figure 12: Illustration of geometric and topological symmetries between correlated linear partitioning systems of a constrained primal normal eigenlocus  $\tau$  and its Wolfe dual  $\psi$ , all of which are created by the strong duality relationships between the algebraic systems of  $\min \Psi(\tau)$  and  $\max \Xi(\psi)$ .

## 6.8 Strong Duality Relationships Between a Constrained Primal and a Wolfe Dual Normal Eigenlocus

All of the constrained primal normal eigenaxis components on a strong dual normal eigenlocus  $\tau$  possess magnitudes and directions that jointly determine the constrained eigen-coordinate locations of an unknown normal eigenaxis of a symmetrical set of linear partitioning curves or surfaces in  $\mathbb{R}^d$ . A comprehensive examination of the statistical decision system of a strong dual normal eigenlocus will reveal how this statistical balancing feat is routinely accomplished by solving the inequality constrained optimization problem of Eq. (13). Sections 7 – 12 will identify strong duality relationships between the algebraic systems of the constrained primal and the Wolfe dual normal eigenaxis components of  $\min \Psi(\tau)$  and  $\max \Xi(\psi)$ . Sections 14 and 15 will demonstrate how the eigenloci of the normal eigenaxis components on the constrained primal normal eigenlocus  $\tau$  of  $\min \Psi(\tau)$  are completely specified by the eigenloci of the Wolfe dual

normal eigenaxis components of  $\max \Xi(\psi)$ . Sections 14 and 15 will also identify geometric and statistical properties which are jointly exhibited by all of the normal eigenaxis components on  $\tau$  and its symmetrical Wolfe dual  $\psi$ .

## 6.9 Uniform Geometric and Statistical Properties Jointly Exhibited by Correlated Normal Eigenaxis Components on $\tau$ and $\psi$

The strong duality relationships between the constrained primal normal eigenlocus of  $\min \Psi(\tau)$  and the Wolfe dual normal eigenlocus of  $\max \Xi(\psi)$  ensure that correlated normal eigenaxis components on  $\tau$  and  $\psi$  exhibit symmetrical geometric and statistical properties. Sections 14 and 15 will demonstrate how the geometric locations of the Wolfe dual normal eigenaxis components are symmetrically correlated to the geometric locations of their constrained primal counterparts, such that all of the constrained primal normal eigenaxis components on  $\tau$  determine principal locations of large covariance. Sections 14 and 15 will also demonstrate that correlated normal eigenaxis components on  $\tau$  and  $\psi$  exhibit directional symmetry. Sections 14 and 15 will examine how eigen-balanced, symmetrical relationships between all of the normal eigenaxis components on  $\tau$  and  $\psi$  determine suitable magnitudes and geometric locations for each of the constrained primal and Wolfe dual normal eigenaxis components. Thereby, it will be shown that the eigenlocus of each constrained primal normal eigenaxis component on  $\tau$  is jointly delineated by the eigenloci of a constrained primal and a Wolfe dual normal eigenaxis component, both of which symmetrically encode magnitudes and directions of large covariance in  $\mathbb{R}^d$  and  $\mathbb{R}^N$  respectively. Thereby, it will be demonstrated that the geometric and statistical properties which are jointly exhibited by  $\tau$  and  $\psi$  involve similarities in magnitudes and directions of correlated constrained primal and Wolfe dual normal eigenaxis components, all of which determine elegant correlations between the total allowed eigenenergies of  $\tau$  and  $\psi$  and the statistical equilibrium point of  $\tau$ . Accordingly, it will be shown that  $\tau$  and  $\psi$  delineate interconnected, dual geometric architectures of symmetrical linear partitions, which jointly determine probabilistic linear discriminant functions. Moreover, it will be shown that the regularized, data-driven geometric architectures, which are jointly delineated by  $\psi$  and  $\tau$ , determine statistical decision systems that provide a robust means for recognizing unknown objects.

## 6.10 Fundamental Relationships Between Joint Statistical Estimates of $\tau$ and $\psi$

It is claimed that the fundamental geometric and statistical property of a strong dual normal eigenlocus  $\tau$  is its total allowed eigenenergy. Furthermore, it is claimed that  $\tau$  exhibits a critical minimum eigenenergy  $\|\tau\|_{\min_e}^2$  which effectively characterizes the geometric configuration of a linear decision boundary and the widths of its decision borders. It is also claimed that  $\tau$  satisfies a linear decision

boundary and its decision borders in terms of its critical minimum eigenenergy.

Given the strong duality relationships between the joint statistical estimates of  $\tau$  and  $\psi$ , it follows that a constrained primal normal eigenlocus  $\tau$  of  $\min \Psi(\tau)$  exhibits a statistical equilibrium point which is symmetrically related to and determined by the eigenlocus of its Wolfe dual  $\psi$  of  $\max \Xi(\psi)$ . Therefore, the total allowed eigenenergies  $\|\tau\|_{\min_c}^2$  and  $\|\psi\|_{\min_c}^2$  of  $\tau$  and  $\psi$  are symmetrically related to each other

$$\|\tau\|_{\min_c}^2 \cong \|\psi\|_{\min_c}^2,$$

in a manner that determines the critical minimum eigenenergy  $\|\tau\|_{\min_c}^2$  of  $\tau$ . Section 17 will develop algebraic and statistical expressions that describe symmetrical relationships between the total allowed eigenenergies  $\|\tau\|_{\min_c}^2$  of  $\tau$  and the magnitudes or lengths of the Wolfe dual normal eigenaxis components on  $\psi$ . Section 17 will develop an identity for which the total allowed eigenenergies  $\|\tau\|_{\min_c}^2 \cong \|\tau_1 - \tau_2\|_{\min_c}^2$  of a strong dual normal eigenlocus  $\tau = \tau_1 - \tau_2$  satisfy the law of cosines in a surprisingly elegant and symmetric manner. Thereby, Section 17 will show that the total allowed eigenenergies  $\|\tau_1 - \tau_2\|_{\min_c}^2$  of  $\tau$  are consistent with the conservation of energy. Section 18 will show that the squares  $\|\psi_{1i*} \mathbf{x}_{1i*}\|_{\min_c}^2$  and  $\|\psi_{2i*} \mathbf{x}_{2i*}\|_{\min_c}^2$  of the constrained primal normal eigenaxis components  $\psi_{1i*} \mathbf{x}_{1i*}$  and  $\psi_{2i*} \mathbf{x}_{2i*}$  determine the probabilities of finding extreme data points  $\mathbf{x}_{1i*}$  and  $\mathbf{x}_{2i*}$  in particular regions of  $\mathbb{R}^d$ , where  $\|\psi_{1i*} \mathbf{x}_{1i*}\|_{\min_c}^2$  is the total allowed eigenenergy of  $\psi_{1i*} \mathbf{x}_{1i*}$  and  $\|\psi_{2i*} \mathbf{x}_{2i*}\|_{\min_c}^2$  is the total allowed eigenenergy of  $\psi_{2i*} \mathbf{x}_{2i*}$ . All of these results will be used to demonstrate how the strong duality relationships between  $\tau$  and  $\psi$  enable joint statistical estimates of the constrained eigen-coordinate locations of an unknown normal eigenaxis  $\mathbf{v}$  in  $\mathbb{R}^d$ .

The regularized Wolfe dual for the strong dual normal eigenlocus problem will be derived by means of the Lagrangian described in the next section. Several strong dual normal eigenlocus equations will be introduced and developed, all of which jointly determine a statistical decision system for probabilistic linear classification.

## 7 The Lagrangian of the Primal Normal Eigenlocus

The inequality constrained optimization problem in Eq. (13) is solved by using Lagrange multipliers  $\psi_i \geq 0$  and the Lagrangian:

$$\begin{aligned} L_{\Psi(\tau)}(\tau, \tau_0, \xi, \psi) = & \|\tau\|^2 / 2 \\ & + C/2 \sum_{i=1}^N \xi_i^2 \\ & - \sum_{i=1}^N \psi_i \\ & \times \{y_i (\mathbf{x}_i^T \tau + \tau_0) - 1 + \xi_i\}. \end{aligned} \tag{15}$$

The Karush-Kuhn-Tucker (KKT) constraints on the Lagrangian  $L_{\Psi(\tau)}$  specify a statistical decision system for probabilistic linear discrimination. It will be shown that the constrained Lagrangian functional  $L_{\Psi(\tau)}$  of Eq. (15) returns the minimum number of normal eigenaxis components that are necessary to symmetrically partition a two class feature space. The KKT constraints on  $L_{\Psi(\tau)}$  are summarized below Cristianini and Shawe-Taylor [2000], Scholkopf and Smola [2002].

### 7.1 A Statistical Decision System for Probabilistic Binary Linear Classification

The KKT constraints on the Lagrangian functional  $L_{\Psi(\tau)}$ :

$$\tau - \sum_{i=1}^N \psi_i y_i \mathbf{x}_i = 0, \quad i = 1, \dots, N, \quad (16)$$

$$\sum_{i=1}^N \psi_i y_i = 0, \quad i = 1, \dots, N, \quad (17)$$

$$C \sum_{i=1}^N \xi_i - \sum_{i=1}^N \psi_i = 0, \quad (18)$$

$$y_i (\mathbf{x}_i^T \tau + \tau_0) - 1 + \xi_i \geq 0, \quad i = 1, \dots, N, \quad (19)$$

$$\psi_i \geq 0, \quad i = 1, \dots, N, \quad (20)$$

$$\psi_i \{y_i (\mathbf{x}_i^T \tau + \tau_0) - 1 + \xi_i\} \geq 0, \quad i = 1, \dots, N, \quad (21)$$

determine a statistical discriminant function

$$D(\mathbf{x}) = \tau^T \mathbf{x} + \tau_0, \quad (22)$$

which satisfies the set of constraints:

$$D_0(\mathbf{x}) = 0,$$

$$D_{+1}(\mathbf{x}) = +1,$$

$$D_{-1}(\mathbf{x}) = -1.$$

It will now be shown that the above set of constraints on Eq. (22) determine three strong dual normal eigenlocus equations of symmetrical linear partitioning curves or surfaces, where each of the points on all three linear loci reference  $\tau$ . Returning to Eq. (4), recall that the locus equation of a normal eigenaxis  $\mathbf{v}$  can be written as:

$$\frac{\mathbf{x}^T \mathbf{v}}{\|\mathbf{v}\|} = \|\mathbf{v}\|,$$

where the normal eigenaxis  $\mathbf{v}/\|\mathbf{v}\|$  has length 1 and points in the direction of the principal eigenvector  $\mathbf{v}$ , such that  $\|\mathbf{v}\|$  is the distance of a specified line, plane, or hyperplane to the origin. Any point  $\mathbf{x}$  that satisfies the above locus equation is on the linear locus of points specified by  $\mathbf{v}$ , where all of the points  $\mathbf{x}$  on the linear locus reference  $\mathbf{v}$ .



Equation (4) and the set of constraints satisfied by the discriminant function  $D(\mathbf{x})$  of Eq. (22) are now used to obtain the set of strong dual normal eigenlocus equations that delineate a linear decision boundary and its bilaterally symmetrical linear decision borders.

### 7.1.1 Eigenlocus Equation of the Linear Decision Boundary

Using Eq. (4) and assuming that  $D(\mathbf{x}) = 0$ , the discriminant function

$$D(\mathbf{x}) = \tau^T \mathbf{x} + \tau_0,$$

can be rewritten as:

$$\frac{\mathbf{x}^T \tau}{\|\tau\|} = -\frac{\tau_0}{\|\tau\|}, \quad (23)$$

where  $\frac{|\tau_0|}{\|\tau\|}$  is the distance of a linear decision boundary  $D_0(\mathbf{x})$  to the origin. Any point  $\mathbf{x}$  that satisfies Eq. (23) is on the linear decision boundary  $D_0(\mathbf{x})$ . All of the points  $\mathbf{x}$  on  $D_0(\mathbf{x})$  reference  $\tau$ . It has been demonstrated by analyses and simulation studies that the linear decision boundary of Eq. (23) optimally partitions the normally distributed training data described by Eq. (97) Reeves [2009].

### 7.1.2 Eigenlocus Equation of the $D_{+1}(\mathbf{x})$ Decision Border

Using Eq. (4) and assuming that  $D(\mathbf{x}) = 1$ , the discriminant function of Eq. (22) can be rewritten as:

$$\frac{\mathbf{x}^T \tau}{\|\tau\|} = -\frac{\tau_0}{\|\tau\|} + \frac{1}{\|\tau\|}, \quad (24)$$

where  $\frac{|1-\tau_0|}{\|\tau\|}$  is the distance of the linear decision border  $D_{+1}(\mathbf{x})$  to the origin. Any point  $\mathbf{x}$  that satisfies Eq. (24) is on the linear decision border  $D_{+1}(\mathbf{x})$ . All of the points  $\mathbf{x}$  on  $D_{+1}(\mathbf{x})$  reference  $\tau$ .

### 7.1.3 Eigenlocus Equation of the $D_{-1}(\mathbf{x})$ Decision Border

Using Eq. (4) and assuming that  $D(\mathbf{x}) = -1$ , the discriminant function of Eq. (22) can be rewritten as:

$$\frac{\mathbf{x}^T \tau}{\|\tau\|} = -\frac{\tau_0}{\|\tau\|} - \frac{1}{\|\tau\|}, \quad (25)$$

where  $\frac{|-1-\tau_0|}{\|\tau\|}$  is the distance of the linear decision border  $D_{-1}(\mathbf{x})$  to the origin. Any point  $\mathbf{x}$  that satisfies Eq. (25) is on the linear decision border  $D_{-1}(\mathbf{x})$ . All of the points  $\mathbf{x}$  on  $D_{-1}(\mathbf{x})$  reference  $\tau$ .

It is concluded that the constrained discriminant function  $D(\mathbf{x})$  of Eq. (22) determines three, symmetrical linear partitioning curves or surfaces, where all of the points on  $D_0(\mathbf{x})$ ,  $D_{+1}(\mathbf{x})$ , and  $D_{-1}(\mathbf{x})$  exclusively reference  $\tau$ . The eigenlocus equations of the linear decision borders are now used to obtain an algebraic expression for the distance between the linear decision borders.

### Distance Between the Linear Decision Borders

Using Eqs (24) and (25), the distance between the linear decision borders  $D_{+1}(\mathbf{x})$  and  $D_{-1}(\mathbf{x})$ :

$$\begin{aligned} D_{(D_{+1}(\mathbf{x})-D_{-1}(\mathbf{x}))} &= \left( -\frac{\tau_0}{\|\tau\|} + \frac{1}{\|\tau\|} \right) \\ &\quad - \left( -\frac{\tau_0}{\|\tau\|} - \frac{1}{\|\tau\|} \right), \\ &= \frac{2}{\|\tau\|}, \end{aligned} \tag{26}$$

is inversely proportional to the length of  $\tau$ . It is concluded that the distance between the linear decision borders is regulated by the term  $\frac{2}{\|\tau\|}$ , which is proportional to the inverted length of a strong dual normal eigenlocus  $\tau$ .

Algebraic expressions are now obtained for the distances between the linear decision boundary and the linear decision borders.

### Distances Between the Linear Decision Boundary and its Borders

Using Eqs (23) and (24), the distance between the linear decision border  $D_{+1}(\mathbf{x})$  and the linear decision boundary  $D_0(\mathbf{x})$  is  $\frac{1}{\|\tau\|}$ :

$$\begin{aligned} D_{(D_{+1}(\mathbf{x})-D_0(\mathbf{x}))} &= \left( -\frac{\tau_0}{\|\tau\|} + \frac{1}{\|\tau\|} \right) \\ &\quad - \left( -\frac{\tau_0}{\|\tau\|} \right), \\ &= \frac{1}{\|\tau\|}. \end{aligned} \tag{27}$$

Using Eqs (23) and (25), the distance between the linear decision boundary  $D_0(\mathbf{x})$  and the linear decision border  $D_{-1}(\mathbf{x})$  is also  $\frac{1}{\|\tau\|}$ :

$$\begin{aligned} D_{(D_0(\mathbf{x})-D_{-1}(\mathbf{x}))} &= \left( -\frac{\tau_0}{\|\tau\|} \right) \\ &\quad - \left( -\frac{\tau_0}{\|\tau\|} - \frac{1}{\|\tau\|} \right), \\ &= \frac{1}{\|\tau\|}. \end{aligned} \tag{28}$$

The equivalent distance of  $\frac{1}{\|\tau\|}$  between each linear decision border and the linear decision boundary reveals that the algebraic and geometric source of the bilateral symmetry of the linear decision borders is the constrained strong dual normal eigenlocus  $\tau$ . It is concluded that the equivalent and constant widths of the bipartite, congruent geometric regions delineated by the linear decision boundary of Eq. (23) and the linear decision borders of Eqs (24) and (25) are regulated by the inverted length  $\frac{1}{\|\tau\|}$  of  $\tau$ .

## 7.2 Axis of Symmetry for Bilateral Linear Partitions

Equations (26), (27), and (28) show that a strong dual normal eigenlocus  $\tau$  determines an axis of symmetry that delineates congruent geometric regions between a linear decision boundary and the bilaterally symmetrical decision borders which bound it. Section 9 will demonstrate that the linear decision borders of Eqs (24) and (25) span (1) geometric regions of data distribution overlap for overlapping data distributions, and (2) geometric regions of large covariance between non-overlapping data distributions. Given this assumption, it is remarkable that a strong dual normal eigenlocus  $\tau$  describes regions of data distribution overlap that exhibit symmetrical widths of  $\frac{1}{\|\tau\|}$ . The sections that follow will determine the manner in which this feat is accomplished.

The next section of the paper will begin to identify geometric and statistical properties exhibited by the primal normal eigenlocus represented within the Wolfe dual eigenspace. The statistical representation of the primal normal eigenlocus within the Wolfe dual eigenspace will be shown to specify a highly interconnected set of constrained primal and Wolfe dual normal eigenaxis components, which are organized in a symmetric manner that encodes essential geometric underpinnings and statistical machinery for a statistical decision system. Section 9 will demonstrate that the statistical representation of the primal normal eigenlocus within the Wolfe dual eigenspace (1) determines a regularized, data-driven geometric architecture that encodes a robust likelihood ratio, and (2) delineates an elegant curve and coordinate system that symmetrically partitions any given feature space.

## 8 Statistical Representation of $\tau$ Within the Wolfe Dual Eigenspace

This section of the paper will begin the process of describing the primal normal eigenlocus within the Wolfe dual eigenspace. Accordingly, the Lagrangian  $L_{\Psi(\tau)}$  is minimized with respect to the primal variables  $\tau$ ,  $\tau_0$ , and  $\xi_i$ , and is maximized with respect to the dual variables  $\psi_i$  Cristianini and Shawe-Taylor [2000], Scholkopf and Smola [2002]. The extrema obtained by representing the primal normal eigenlocus within the Wolfe dual eigenspace are summarized below.

### 8.1 The Constrained Primal Normal Eigenlocus

The KKT constraint of Eq. (20) restricts the length  $\psi_i$  of any Wolfe dual normal eigenaxis component  $\psi_i \vec{e}_i$  on  $\psi$  to either satisfy or exceed zero:  $\psi_i \geq 0$ . Any Wolfe dual normal eigenaxis component  $\psi_i \vec{e}_i$  which has the length  $\psi_i = 0$  is *not on* the Wolfe dual normal eigenlocus  $\psi$ . It follows that the constrained primal normal eigenaxis component  $\psi_i \mathbf{x}_i$  which has the length  $\|\psi_i \mathbf{x}_i\| = 0$  is *not on* the constrained primal normal eigenlocus  $\tau$ . The KKT constraints of Eqs (16) and (20) jointly determine the primal normal eigenlocus  $\tau$  within the Wolfe dual eigenspace, so that an estimate for  $\tau$  satisfies the following strong dual normal

eigenlocus equation:

$$\tau = \sum_{i=1}^N y_i \psi_i \mathbf{x}_i, \quad (29)$$

where the  $y_i$  terms are training set labels (if  $\mathbf{x}_i$  is a member of pattern class one, assign  $y_i = 1$ ; otherwise, assign  $y_i = -1$ ) and the magnitude  $\psi_i$  of each Wolfe dual normal eigenaxis component  $\psi_i \vec{\mathbf{e}}_i$  is greater than or equal to zero:  $\psi_i \geq 0$ . Training points  $\mathbf{x}_i$  which are correlated with Wolfe dual normal eigenaxes  $\psi_i \vec{\mathbf{e}}_i$  that have non-zero magnitudes or lengths  $\psi_i > 0$  are termed *extreme* training vectors. Accordingly, extreme training vectors are essentially unconstrained primal normal eigenaxis components. Extreme training points are innermost data points of large covariance that are located between overlapping or non-overlapping data distributions. Given these assumptions, Eq. (29) determines a dual statistical eigenlocus of normal eigenaxis components formed by eigen-scaled extreme training points, all of which encode principal locations of large covariance. The location properties of extreme training points are defined next.

## Location Properties of Extreme Training Points

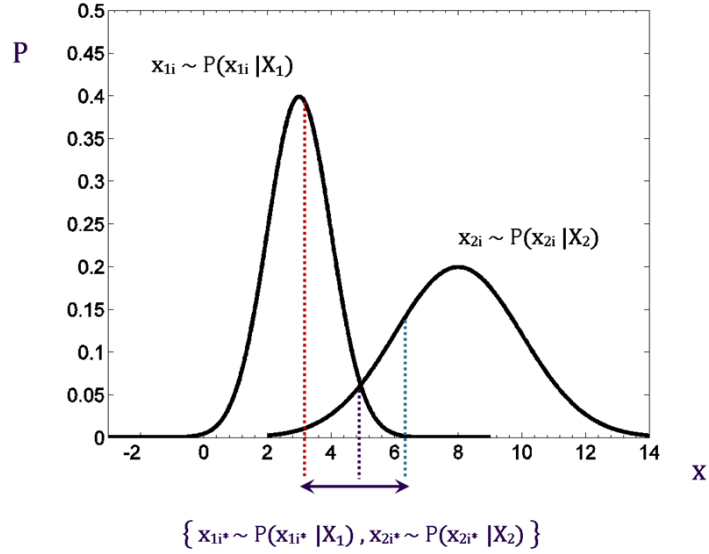
Take a collection of training data drawn from any two probability distributions. An extreme training point is defined to be a data point which exhibits a high variability of geometric location, that is, possesses a large covariance, such that it is located (1) relatively far from its distribution mean, (2) relatively close to the mean of the other distribution, and (2) relatively close to other extreme points. Accordingly, an extreme data point is located somewhere between a pair of overlapping or non-overlapping data distributions. Given the location properties exhibited by the geometric locus of an extreme data point, it follows that a set of extreme vectors determine principal directions of large covariance for a given collection of training data. Likewise, the geometric loci of a set of extreme vectors span a geometric region of large covariance. Therefore, a set of extreme training points span a geometric region of large covariance that is located between two distributions of training data. It follows that the geometric loci of any given set of extreme vectors span a particular region of  $\mathbb{R}^d$ .

It will now be argued that extreme training vectors are unconstrained primal normal eigenaxis components used to form  $\tau$ . Section 18 will demonstrate that each constrained primal normal eigenaxis component describes the probability of finding an extreme data point in a particular region of  $\mathbb{R}^d$ . Thereby, Section 18 will show that the integrated set of constrained primal normal eigenaxis components on  $\tau_1$  and  $\tau_2$ , i.e., on  $\tau = \tau_1 - \tau_2$ , describes the probabilities of finding each of the extreme data points in particular regions of  $\mathbb{R}^d$ , where all of the extreme data points are located in regions of large covariance between either overlapping or non-overlapping data distributions.

The location properties of extreme data points for overlapping and non-overlapping data distributions are defined next.

### 8.1.1 Extreme Data Points of Overlapping Data Distributions

For overlapping data distributions, the geometric loci of the extreme data points from each pattern class are distributed within bipartite, joint geometric regions of large covariance, both of which span the region of data distribution overlap. Figure 13 depicts bipartite, joint geometric regions of large variance that are located between two overlapping data distributions.



**Bipartite Joint Regions of Extreme Data Points  $x_{1i*}$  and  $x_{2i*}$**

FIGURE 13. Illustration of extreme data points, denoted by  $x_{1i*}$  and  $x_{2i*}$ , which are located in bipartite, joint geometric regions of large variance that are positioned between two overlapping data distributions.

### 8.1.2 Extreme Data Points of Non-overlapping Data Distributions

For non-overlapping data distributions, the geometric loci of the extreme data points are distributed within bipartite, disjoint geometric regions of large covariance, i.e., separate tail regions, that are located between the data distributions. Because tail regions of distributions determine intervals of low probability, it follows that relatively few extreme data points are located within tail regions. Therefore, relatively few extreme data points are located between non-overlapping data distributions. Figure 14 illustrates how a small number of extreme data points are located within the tail regions of non-overlapping Gaussian data distributions.

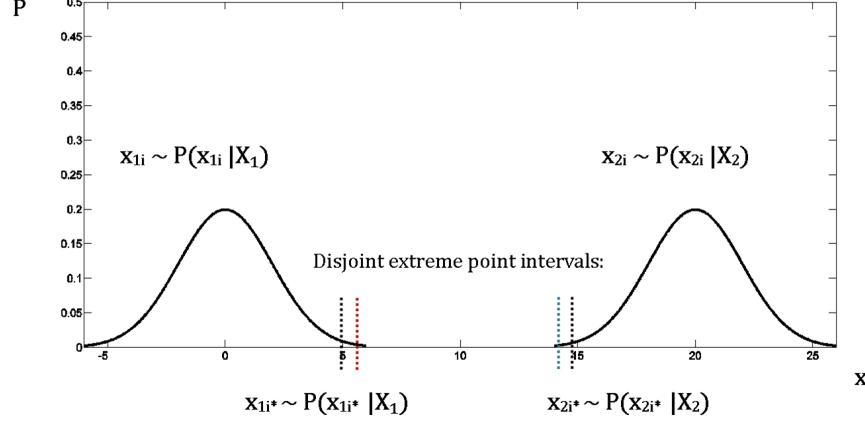


FIGURE 14. Illustration of how relatively few extreme data points, denoted by  $\mathbf{x}_{1i*}$  and  $\mathbf{x}_{2i*}$ , are located in the tail regions of non-overlapping Gaussian data distributions.

Sections 14 and 15 will demonstrate that each Wolfe dual normal eigenaxis component encodes an eigen-balanced, unidirectional pointwise covariance estimate for an extreme data point, which specifies an eigen-scale for that extreme training vector. The next section will consider how a constrained primal normal eigenlocus  $\tau$  is formed by a pair of strong dual, i.e., constrained primal, normal eigenlocus components.

## 8.2 The Pair of Strong Dual Normal Eigenlocus Components

All of the constrained primal normal eigenaxis components on a strong dual normal eigenlocus  $\tau$  are labeled, eigen-scaled extreme training points in  $\mathbb{R}^d$ . Denote the eigen-scaled extreme training vectors that belong to pattern classes one and two by  $\psi_{1i*} \mathbf{x}_{1i*}$  and  $\psi_{2i*} \mathbf{x}_{2i*}$ , with eigen-scales  $\psi_{1i*}$  and  $\psi_{2i*}$ , extreme training vectors  $\mathbf{x}_{1i*}$  and  $\mathbf{x}_{2i*}$ , and training labels  $y_i = 1$  and  $y_i = -1$  respectively. Let there be  $l_1$  eigen-scaled extreme training points  $\{\psi_{1i*} \mathbf{x}_{1i*}\}_{i=1}^{l_1}$  and  $l_2$  eigen-scaled extreme training points  $\{\psi_{2i*} \mathbf{x}_{2i*}\}_{i=1}^{l_2}$ .

Given Eq. (29) and the assumptions outlined above, it follows that an estimate for a strong dual normal eigenlocus  $\tau$  is based on the vector difference between a pair of constrained primal normal eigenlocus components:

$$\begin{aligned} \tau &= \sum_{i=1}^{l_1} \psi_{1i*} \mathbf{x}_{1i*} - \sum_{i=1}^{l_2} \psi_{2i*} \mathbf{x}_{2i*}, \\ &= \tau_1 - \tau_2, \end{aligned} \quad (30)$$

where the constrained primal normal eigenlocus components  $\sum_{i=1}^{l_1} \psi_{1i*} \mathbf{x}_{1i*}$  and  $\sum_{i=1}^{l_2} \psi_{2i*} \mathbf{x}_{2i*}$  are denoted by  $\tau_1$  and  $\tau_2$ . The eigen-scaled extreme training

points  $\{\psi_{11*}\mathbf{x}_{1i*}\}_{i=1}^{l_1}$  and  $\{\psi_{21*}\mathbf{x}_{2i*}\}_{i=1}^{l_2}$  on  $\tau_1$  and  $\tau_2$  determine the eigenloci of  $\tau_1$  and  $\tau_2$ , and thereby determine the eigenlocus of  $\tau = \tau_1 - \tau_2$ . Figure 15 depicts how the geometric configurations of the  $\tau_1$  and  $\tau_2$  strong dual normal eigenlocus components of  $\tau$  effectively determine the geometric configuration of  $\tau$ .

### The Primal Normal Eigenlocus Represented in the Wolfe Dual Eigenspace

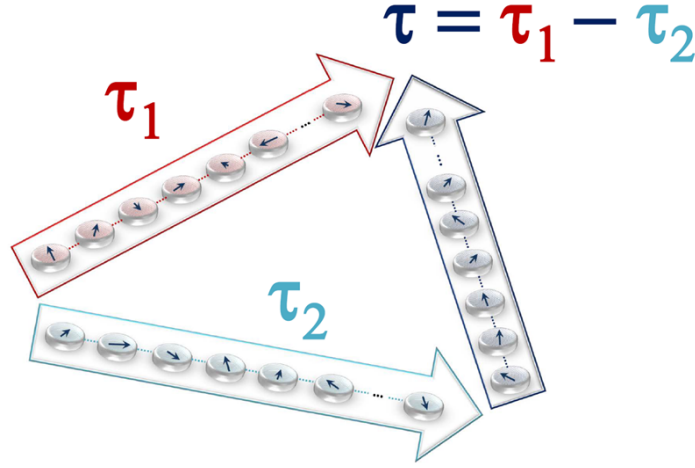


Figure 15: Illustration of how the primal normal eigenlocus  $\tau$  represented in the Wolfe dual eigenspace is formed by the vector difference  $\tau_1 - \tau_2$  between a pair of constrained primal normal eigenlocus components  $\tau_1$  and  $\tau_2$ . The eigen-scaled extreme points on  $\tau_1$ ,  $\tau_2$ , and  $\tau$  are depicted by variable length arrows pointing in various directions, which illustrate eigen-scaled extreme training vectors that possess unchanged directions and eigen-balanced lengths.

It will now be demonstrated how the eigenloci of the constrained primal normal eigenlocus components  $\tau_1$  and  $\tau_2$  regulate the geometric width, i.e., the breadth of the geometric region, between the linear decision borders  $D_1(\mathbf{x})$  and  $D_{-1}(\mathbf{x})$ .

## 9 Width Regulation of Linear Decision Regions

Substitution of the expression for  $\tau$  in Eq. (30) into Eq. (26) provides a new expression for the width of the geometric region between the linear decision borders  $D_1(\mathbf{x})$  and  $D_{-1}(\mathbf{x})$ :

$$D_{(D_1(\mathbf{x})-D_{-1}(\mathbf{x}))} = \frac{2}{\|\tau_1 - \tau_2\|}, \quad (31)$$

where the constrained width of the geometric region between the linear decision borders is inversely proportional to the magnitude of the vector difference of  $\tau_1$  and  $\tau_2$ . Equation (31) shows that the span of the geometric region between the linear decision borders is regulated by the magnitudes and the directions of the constrained primal normal eigenlocus components  $\tau_1$  and  $\tau_2$ .

The eigenloci of the constrained primal normal eigenlocus components  $\tau_1$  and  $\tau_2$  also regulate the span of the congruent geometric regions between the linear decision boundary  $D_0(\mathbf{x})$  and the linear decision borders  $D_{+1}(\mathbf{x})$  and  $D_{-1}(\mathbf{x})$ . Substitution of the expression for  $\tau$  in Eq. (30) into Eq. (27) provides a new expression for the width of the geometric region between the linear decision border  $D_{+1}(\mathbf{x})$  and the linear decision boundary  $D_0(\mathbf{x})$ :

$$\begin{aligned} D_{(D_{+1}(\mathbf{x})-D_0(\mathbf{x}))} &= \left( -\frac{\tau_0}{\|\tau_1 - \tau_2\|} + \frac{1}{\|\tau_1 - \tau_2\|} \right) \\ &\quad - \left( -\frac{\tau_0}{\|\tau_1 - \tau_2\|} \right), \\ &= \frac{1}{\|\tau_1 - \tau_2\|}, \end{aligned} \quad (32)$$

where the width of the geometric region between  $D_0(\mathbf{x})$  and  $D_{+1}(\mathbf{x})$  satisfies  $\frac{1}{\|\tau_1 - \tau_2\|}$ .

Likewise, the span of the geometric region between the linear decision boundary  $D_0(\mathbf{x})$  and the linear decision border  $D_{-1}(\mathbf{x})$ :

$$\begin{aligned} D_{(D_0(\mathbf{x})-D_{-1}(\mathbf{x}))} &= \left( -\frac{\tau_0}{\|\tau_1 - \tau_2\|} \right) \\ &\quad - \left( -\frac{\tau_0}{\|\tau_1 - \tau_2\|} - \frac{1}{\|\tau_1 - \tau_2\|} \right), \\ &= \frac{1}{\|\tau_1 - \tau_2\|}, \end{aligned} \quad (33)$$

also satisfies  $\frac{1}{\|\tau_1 - \tau_2\|}$ .

It is concluded that the width of the bipartite, congruent geometric regions between the linear decision boundary and the linear decision borders is inversely proportional to the magnitude of the vector difference of  $\tau_1$  and  $\tau_2$ :

$$\frac{1}{\|\tau_1 - \tau_2\|},$$

which indicates that the balanced geometric widths of the symmetrical decision regions of the constrained Lagrangian of Eq. (15) are regulated by the magnitudes and the directions of the constrained primal normal eigenlocus components  $\tau_1$  and  $\tau_2$ .



## 9.1 Bipartite Symmetric Partitions of Large Covariance Regions

It will now be argued that the bipartite, symmetrical decision regions delineated by the linear decision boundary of Eq. (23) and the linear decision borders of Eqs (24) and (25), describe symmetric regions of large covariance for both overlapping and non-overlapping data distributions. Recall that, by definition, extreme training points exhibit a high variability of geometric location and therefore possess a large covariance, such that the geometric locations of a set of extreme training points span a geometric region of large covariance that is located between two distributions of training data. Therefore, by definition, the width of any large covariance geometric region depends on the geometric loci of the extreme vectors of the distributions.

### Assumptions

At this stage of the analysis, it is necessary to develop more of the geometric underpinnings and statistical machinery that is produced by the constrained Lagrangian of Eq. (15). For this reason, several significant results will be assumed that will be substantiated later on. Section 18 will show that the geometric configuration of the linear decision boundary, and the widths of the bipartite, congruent geometric regions located between the linear decision boundary of Eq. (23) and the linear decision borders of Eqs (24) and (25), are regulated by the probability of finding extreme data points in particular regions of  $\mathbb{R}^d$ .

For now, it is assumed that the integrated set of constrained primal normal eigenaxis components on  $\tau = \tau_1 - \tau_2$  describes the probabilities of finding the extreme data points in particular regions of  $\mathbb{R}^d$ , where all of the extreme data points are located in regions of large covariance between either overlapping or non-overlapping data distributions. Thereby, it is assumed that the constrained primal normal eigenaxis components on  $\tau_1$  and  $\tau_2$  describe disjoint tail regions between non-overlapping data distributions, and bipartite, joint geometric regions of large covariance between overlapping data distributions. The next section will examine strong dual normal eigenlocus transforms for non-overlapping data distributions.

## 9.2 Strong Dual Normal Eigenlocus Transforms for Non-overlapping Data Distributions

It will now be argued that the linear decision boundary of Eq. (23) and the linear decision borders of Eqs (24) and (25) delineate symmetric, non-overlapping regions of large covariance for any two non-overlapping data distributions.

Take a collection of training data generated by any two non-overlapping probability distributions, where all of the extreme data points are located within the bipartite, disjoint tail regions of the distributions. Given these assumptions and Eq. (30), it follows that the strong dual normal eigenlocus components  $\tau_1$  and  $\tau_2$  on  $\tau = \tau_1 - \tau_2$  are formed by relatively few eigen-scaled extreme data

points, i.e.,  $\tau_1 = \sum_{i=1}^{l_1} \psi_{1_{i*}} \mathbf{x}_{1_{i*}}$  and  $\tau_2 = \sum_{i=1}^{l_2} \psi_{2_{i*}} \mathbf{x}_{2_{i*}}$ , all of which describe the probabilities of finding the extreme data points in the tail regions of two data distributions in  $\mathbb{R}^d$ . Given Eqs (30) and (31), it follows that the width  $\frac{2}{\|\tau_1 - \tau_2\|}$  of the geometric region located between the linear decision borders of Eqs (24) and (25) is regulated by the eigen-transformed locations of the extreme training points on  $\tau_1 - \tau_2$ , where each eigen-transformed location of an extreme data point describes the probability of finding the extreme data point in the tail region of a data distribution in  $\mathbb{R}^d$ . Given Eqs (30), (32), and (33), it follows that the equivalent widths  $\frac{1}{\|\tau_1 - \tau_2\|}$  of the congruent geometric regions, which are located between the linear decision boundary of Eq. (23) and the linear decision borders of Eqs (24) and (25), are regulated by the eigen-transformed locations of the extreme training points on  $\tau_1 - \tau_2$ .

Given the above assumptions and chain of arguments, it is concluded that the bipartite, congruent geometric regions located between the linear decision boundary of Eq. (23) and the linear decision borders of Eqs (24) and (25) delineate bipartite, congruent, non-overlapping geometric regions of large covariance for non-overlapping data distributions. It is also concluded that the linear decision borders of Eqs (24) and (25) delineate a geometric region of large covariance that spans a geometric region between the tails of two data distributions.

### 9.3 Beyond Classical Interpolation Methods

It is well known that the components of the extreme training vectors from each of the pattern classes satisfy their respective decision borders for non-overlapping data distributions Cortes and Vapnik [1995], Cristianini and Shawe-Taylor [2000], Hastie et al. [2001], Scholkopf and Smola [2002]. However, the mathematical machinery behind classical interpolation or regression methods provides no real insight into what is actually going on.

Given non-overlapping data distributions and two extreme training points  $\mathbf{x}_{1*}$  and  $\mathbf{x}_{2*}$ , it can be shown that the symmetrical eigen-scale  $\psi_{1*} = \psi_{2*}$  for each extreme training vector is the reciprocal of the inner product of the vector difference  $\mathbf{x}_{1*} - \mathbf{x}_{2*}$

$$\psi_{1_{i*}} = \psi_{2_{i*}} = \frac{2}{\|\mathbf{x}_{1_{i*}} - \mathbf{x}_{2_{i*}}\|^2},$$

of  $\mathbf{x}_{1*}$  and  $\mathbf{x}_{2*}$ , which indicates that the magnitudes of the Wolfe dual normal eigenaxis components involve second-order distance statistics between the locations of the extreme points.

Using the above expression and Eq. (31), it follows that the width of the geometric region between the linear decision borders  $D_1(\mathbf{x})$  and  $D_{-1}(\mathbf{x})$  is determined by the eigen-transformed locations of the extreme training points

$\mathbf{x}_{1*}$  and  $\mathbf{x}_{2*}$

$$\begin{aligned} \frac{2}{\|\tau_1 - \tau_2\|} &= 2 \left\| \frac{2\mathbf{x}_{1*}}{\|\mathbf{x}_{1*} - \mathbf{x}_{2*}\|^2} - \frac{2\mathbf{x}_{2*}}{\|\mathbf{x}_{1*} - \mathbf{x}_{2*}\|^2} \right\|^{-1}, \\ &= 2 \left\| \frac{2(\mathbf{x}_{1*} - \mathbf{x}_{2*})}{\|\mathbf{x}_{1*}\|^2 + \|\mathbf{x}_{2*}\|^2 - 2\|\mathbf{x}_{1*}\|\|\mathbf{x}_{2*}\|\cos\theta_{\mathbf{x}_{1*}\mathbf{x}_{2*}}} \right\|^{-1}, \end{aligned}$$

which reduces to

$$\begin{aligned} \frac{2}{\|\tau_1 - \tau_2\|} &= 2 \left\| \frac{2}{(\mathbf{x}_{1*} - \mathbf{x}_{2*})} \right\|^{-1}, \\ &= \|\mathbf{x}_{1*} - \mathbf{x}_{2*}\|. \end{aligned}$$

Returning to Eq. (11) and Fig. 9 in Section 5, it follows that the span of the geometric region between the linear decision borders  $D_1(\mathbf{x})$  and  $D_{-1}(\mathbf{x})$  is determined by the distance between the geometric loci of the extreme training points  $\mathbf{x}_{1*}$  and  $\mathbf{x}_{2*}$ . Thereby, the geometric loci of the extreme training vectors from each of the pattern classes satisfy their respective decision borders.

Using the expression for the symmetrical eigen-scale  $\psi_{1*} = \psi_{2*}$  and Eqs (32) and (33), it follows that the symmetrical widths of the non-overlapping regions of large covariance located between the linear decision border  $D_1(\mathbf{x})$  or  $D_{-1}(\mathbf{x})$  and the linear decision boundary  $D_0(\mathbf{x})$  are determined by equally proportioned eigen-transformed locations of  $\mathbf{x}_{1*}$  and  $\mathbf{x}_{2*}$ .

$$\begin{aligned} \frac{1}{\|\tau_1 - \tau_2\|} &= \left\| \frac{2(\mathbf{x}_{1*} - \mathbf{x}_{2*})}{\|\mathbf{x}_{1*}\|^2 + \|\mathbf{x}_{2*}\|^2 - 2\|\mathbf{x}_{1*}\|\|\mathbf{x}_{2*}\|\cos\theta_{\mathbf{x}_{1*}\mathbf{x}_{2*}}} \right\|^{-1} \\ &= \frac{1}{2} \|\mathbf{x}_{1*} - \mathbf{x}_{2*}\|, \end{aligned}$$

which indicates that the symmetrical widths of the decision regions are determined by the half the distance between the geometric loci of the extreme training points  $\mathbf{x}_{1*}$  and  $\mathbf{x}_{2*}$ .

For non-overlapping data distributions, simulation studies show that the extreme training points from each of the pattern classes lie on their respective decision borders. Figure 16 illustrates how the constrained discriminant function of Eq. (22) delineates bipartite, congruent, non-overlapping geometric regions of large covariance for two non-overlapping Gaussian data distributions. Figure 16 also shows that the symmetrical widths of the non-overlapping regions of large covariance are determined by the locations of the extreme vectors of the data distributions.

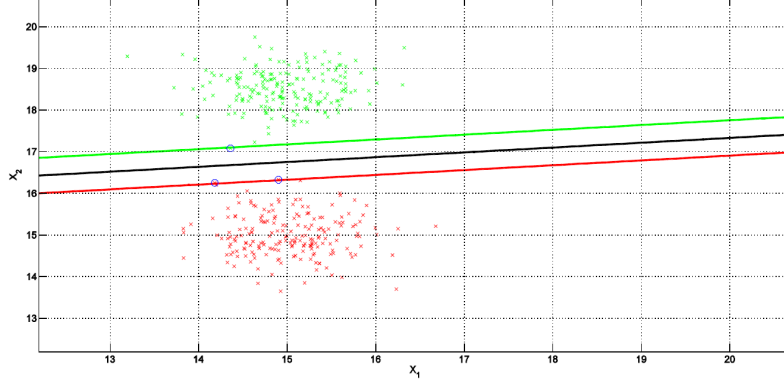


Figure 16: Illustration of the geometric configurations of a linear decision boundary and bilaterally symmetrical linear decision borders, for two non-overlapping data distributions, which are determined by eigen-transformed locations of three extreme training points, all of which lie on their respective decision borders. Each extreme training point is enclosed in a blue circle.

In general, for any given pair of non-overlapping data distributions, the symmetrical widths of the bipartite, congruent, non-overlapping geometric regions of large covariance delineated by Eqs (23), (24), and (25), are a function of eigen-balanced distances between the geometric loci of the extreme points of the data distributions. Figure 17 depicts the symmetrical decision regions of large covariance that are delineated by a constrained strong dual normal eigen-locus discriminate function  $\tau^T \mathbf{x} + \tau_0$  for non-overlapping data distributions.

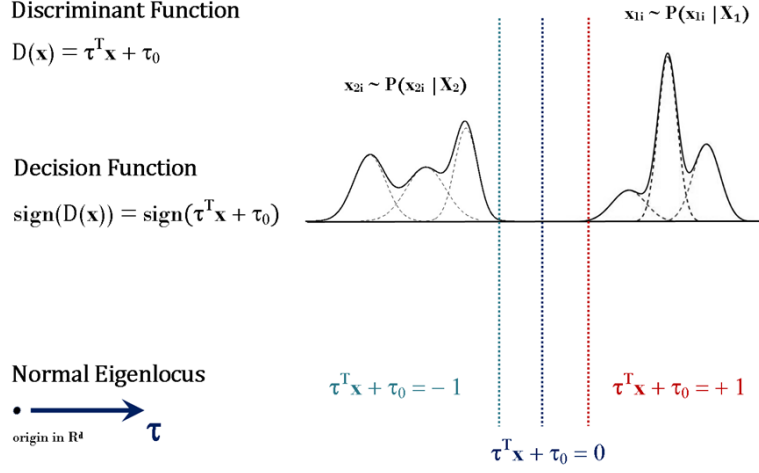


Figure 17: Depiction of the bipartite, symmetrical decision regions of large covariance that are delineated by a constrained strong dual normal eigenlocus discriminant function for non-overlapping data distributions.

The next section will examine strong dual normal eigenlocus transforms for overlapping data distributions.

#### 9.4 Strong Dual Normal Eigenlocus Transforms for Overlapping Data Distributions

It will now be argued that the linear decision boundary of Eq. (23) and the linear decision borders of Eqs (24) and (25) delineate symmetric regions of data distribution overlap for any two overlapping data distributions.

Take a collection of training data generated by any two overlapping probability distributions, where all of the extreme data points are located within the bipartite, joint (overlapping) geometric regions of large covariance that span the region of data distribution overlap. For any given collection of training data, both the number and the locations of the extreme data points are determined by the probability density functions of the training data. Therefore, the geometric shape or configuration of the data distribution overlap is determined by the probability density functions of the training data.

Now take a pair of overlapping data distributions. Let there be  $l_1$  extreme training points  $\{\mathbf{x}_{1i*}\}_{i=1}^{l_1}$  and  $l_2$  extreme training points  $\{\mathbf{x}_{2i*}\}_{i=1}^{l_2}$ . Given the above assumptions and Eq. (30), it follows that the strong dual normal eigenlocus components  $\tau_1$  and  $\tau_2$  on  $\tau = \tau_1 - \tau_2$  are formed by  $l_1 + l_2$  eigen-scaled

extreme data points, i.e.,  $\tau_1 = \sum_{i=1}^{l_1} \psi_{1i*} \mathbf{x}_{1i*}$  and  $\tau_2 = \sum_{i=1}^{l_2} \psi_{2i*} \mathbf{x}_{2i*}$ , all of which jointly describe the bipartite, joint geometric regions of large covariance between the overlapping data distributions. It is assumed that each eigen-scaled extreme data point  $\psi_{1i*} \mathbf{x}_{1i*}$  or  $\psi_{2i*} \mathbf{x}_{2i*}$  describes the probability of finding that extreme data point  $\mathbf{x}_{1i*}$  or  $\mathbf{x}_{2i*}$  within a specific region between the overlapping data distributions.

Given Eqs (30) and (31), it follows that the width  $\frac{2}{\|\tau_1 - \tau_2\|}$  of the geometric region located between the linear decision borders of Eqs (24) and (25) is regulated by the eigen-transformed locations of the extreme training points on  $\tau_1 - \tau_2$ :

$$\frac{2}{\left\| \sum_{i=1}^{l_1} \psi_{1i*} \mathbf{x}_{1i*} - \sum_{i=1}^{l_2} \psi_{2i*} \mathbf{x}_{2i*} \right\|},$$

where each eigen-scaled extreme training point describes the probability of finding the extreme point within a region of data distribution overlap.

It is concluded that the total width  $\frac{2}{\|\tau_1 - \tau_2\|}$  of the geometric region between the linear decision borders delineated by Eq. (31) is regulated by the eigen-transformed locations of the extreme vectors of the data distributions, where the eigen-transformed location of an extreme training point describes the probability of finding that extreme data point within a specific region between the overlapping data distributions. Therefore, it is concluded that the geometric region located between the linear decision borders of Eqs (24) and (25) spans the region of data distribution overlap.

Given Eqs (30), (32), and (33), it follows that the equivalent widths  $\frac{1}{\|\tau_1 - \tau_2\|}$  of the bipartite, congruent geometric regions located between the linear decision boundary of Eq. (23) and the linear decision borders of Eqs (24) and (25), are regulated by the eigen-transformed locations of all of the extreme training points on  $\tau_1 - \tau_2$ :

$$\frac{1}{\left\| \sum_{i=1}^{l_1} \psi_{1i*} \mathbf{x}_{1i*} - \sum_{i=1}^{l_2} \psi_{2i*} \mathbf{x}_{2i*} \right\|}.$$

This implies that the balanced widths  $\frac{1}{\|\tau_1 - \tau_2\|}$  of the congruent geometric regions of distribution overlap delineated by Eqs (32) and (33), are determined by eigen-transformed locations of the extreme vectors of the data distributions, where the geometric loci of the extreme vectors determine the amount of data distribution overlap, and the eigen-transformed locations of the extreme vectors describe the probabilities of finding the extreme data points within specific regions between the overlapping data distributions. It is concluded that the linear decision boundary of Eq. (23) and the linear decision borders of Eqs (24) and (25) delineate bipartite, congruent, large covariance geometric regions of data distribution overlap for any given pair of overlapping data distributions.

It has been demonstrated by simulation studies that the constrained discriminate function  $\tau^T \mathbf{x} + \tau_0$  of Eq. (22) does indeed delineate bipartite, congruent geometric regions of data distribution overlap Reeves [2007], Reeves [2009]. For example, consider Figs. 5 and 7 of Section 3. In general, strong dual normal

eigenlocus decision systems delineate bipartite, symmetrical decision regions for any given pair of overlapping data distributions. Figure 18 depicts the symmetrical decision regions delineated by a constrained discriminant function  $\tau^T \mathbf{x} + \tau_0$  for overlapping data distributions with different covariance structures.

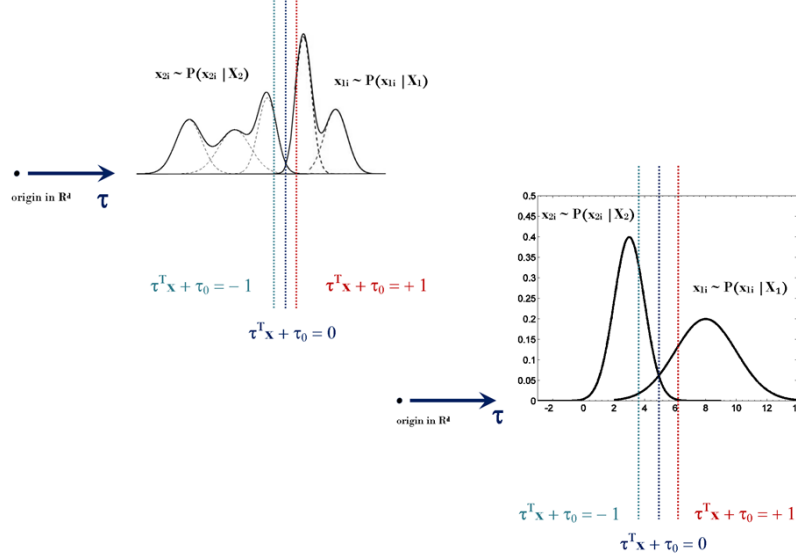


Figure 18: Depiction of the bipartite, symmetrical decision regions delineated by a constrained strong dual normal eigenlocus discriminant function for overlapping data distributions with different covariance structures.

Consider the following example of overlapping Gaussian data distributions with different covariance structures. Figure 19 illustrates the bipartite, symmetrical decision regions delineated by a constrained discriminant function  $\tau^T \mathbf{x} + \tau_0$  for two Gaussian data sets that have the covariance matrices

$$\Sigma_1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \text{ and } \Sigma_2 = \begin{pmatrix} 0.25 & 0 \\ 0 & 5 \end{pmatrix},$$

and the mean vectors  $\mu_1 = (1, 2)^T$  and  $\mu_2 = (0, 2)^T$ . The constrained discriminant function determines a centrally located linear decision boundary that symmetrically partitions the feature space.

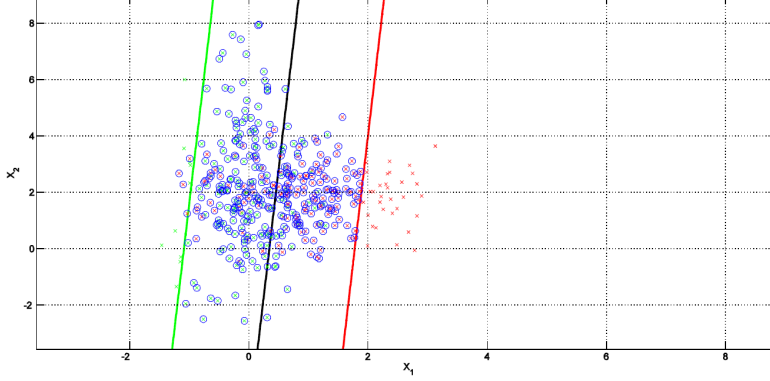


Figure 19: Illustration of the bipartite, symmetrical decision regions delineated by a constrained discriminant function  $\tau^T \mathbf{x} + \tau_0$  for overlapping data distributions with different covariance structures. The centrally located linear decision boundary symmetrically partitions the feature space. The linear decision borders delineate a geometric region of large covariance that spans bipartite, congruent geometric regions of data distribution overlap.

## 9.5 Regularized and Customized Geometric Architectures

Take any given pair of data distributions whose expected values and covariance structures do not vary over time. This paper will demonstrate how strong dual normal eigenlocus transforms produce *customized and regularized geometric architectures that encode robust decision statistics for the binary classification task*.

Furthermore, Figs 15 and 20 illustrate that the data-driven directions and eigen-balanced magnitudes which can be realized by the strong dual normal eigenlocus components  $\tau_1$  and  $\tau_2$  on  $\tau = \tau_1 - \tau_2$ , determine an unlimited number of customized, regularized geometric architectures that can be implemented by the strong dual decision system of Eqs ( 22), (23), (24), and (25).

Figure 20 depicts the regularized, data-driven geometric architecture of a strong dual normal eigenlocus in the Euclidean plane.



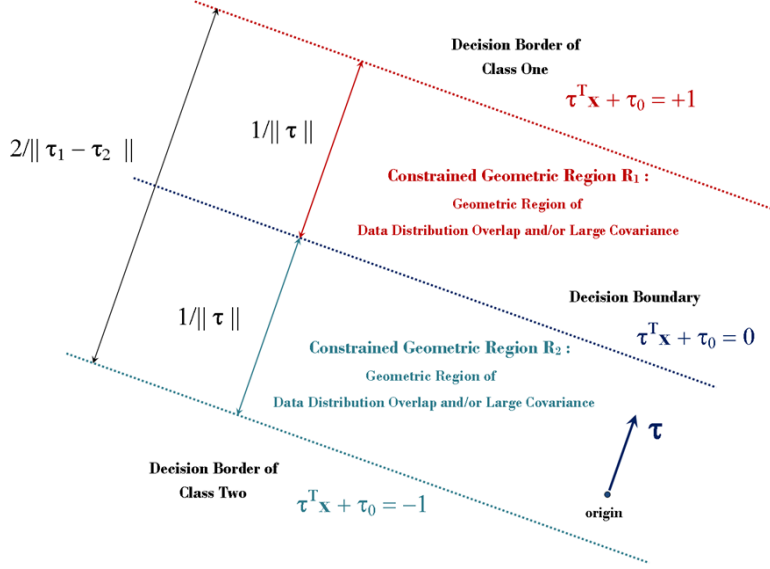


Figure 20: Illustration of a statistical decision system of a strong dual normal eigenlocus in the Euclidean plane. The algebraic system of strong dual normal eigenlocus equations satisfied by  $\tau$ ,  $\tau_1$ ,  $\tau_2$ , and  $\tau_0$  specifies the geometric configurations of  $\tau_1$ ,  $\tau_2$ , and  $\tau$ , which jointly specify the configurations of the constrained geometric regions of large covariance denoted by  $R_1$  and  $R_2$ .

So far, this paper has argued that the KKT constraints on the Lagrangian functional  $L_{\Psi(\tau)}$  jointly specify a set of symmetrical eigen-scales for a set of extreme training points that are located between two given distributions of training data. Later on, Sections 14 and 15 will examine how the eigenlocus of each constrained primal normal eigenaxis component  $\psi_{1i*} \mathbf{x}_{1i*}$  or  $\psi_{2i*} \mathbf{x}_{2i*}$  on  $\tau_1$  or  $\tau_2$  specifies a principal region of high variability, which contributes to the symmetrical partitioning of a region of large covariance located between two data distributions.

Now, consider again the regularized, data-driven geometric architecture depicted in Fig. 20. Given that (1) the total allowed eigenenergy of  $\|\tau\|_{\min_c}^2$  of  $\tau$  satisfies the linear decision boundaries and decision borders depicted in Figs 17, 18, and 20, and that (2) the magnitudes and directions of  $\tau_1$  and  $\tau_2$  regulate the symmetrical configurations of the constrained geometric regions of large covariance depicted in Figs 17, 18, and 20, it is claimed that the total allowed eigenenergies of  $\tau_1$  and  $\tau_2$  must be symmetrically balanced with each other. It will now be argued that the critical minimum eigenenergies of  $\tau_1$  and  $\tau_2$  must be balanced in a symmetric manner.

## 9.6 Balancing the Total Allowed Eigenenergies of $\tau_1$ and

$\tau_2$

Substitution of the expression for the pair of constrained primal normal eigenlocus components  $\tau_1 - \tau_2$  on  $\tau$  in Eq. (30) into the critical minimum eigenenergy constraint for  $\tau$  in Eq. (14) produces an expression

$$\|\tau_1 - \tau_2\|_{\min_c}^2 \cong \left( \|\tau_1\|_{\min_c}^2 + \|\tau_2\|_{\min_c}^2 - \tau_2^T \tau_1 - \tau_1^T \tau_2 \right),$$

which shows that the normal eigenlocus components  $\tau_1$  and  $\tau_2$  on the constrained primal normal eigenlocus  $\tau_1 - \tau_2$  must satisfy the critical minimum eigenenergy constraint

$$\|\tau_1 - \tau_2\|_{\min_c}^2 \cong \|\tau_1\|_{\min_c}^2 + \|\tau_2\|_{\min_c}^2 - 2\tau_1^T \tau_2.$$

It is claimed that the total allowed eigenenergies of  $\tau_1$  and  $\tau_2$  are symmetrically balanced with each other by means of a symmetric equalizer statistic  $\nabla_{eq}$  in relation to a centrally located statistical fulcrum  $f_s$ . Given the assumption that the eigenenergies of  $\tau_1$  and  $\tau_2$  are balanced in a symmetric manner, it is claimed that the total allowed eigenenergy  $\|\tau_1 - \tau_2\|_{\min_c}^2$  of  $\tau$

$$\|\tau_1 - \tau_2\|_{\min_c}^2 \cong \left\{ \|\tau_1\|_{\min_c}^2 - \tau_1^T \tau_2 \right\} + \left\{ \|\tau_2\|_{\min_c}^2 - \tau_2^T \tau_1 \right\},$$

is minimized when the total allowed eigenenergy of  $\tau_1$  satisfies the expression

$$\|\tau_1\|_{\min_c}^2 - \|\tau_1\| \|\tau_2\| \cos \theta_{\tau_1 \tau_2} \cong f_s - \nabla_{eq},$$

and the total allowed eigenenergy of  $\tau_2$  satisfies the expression

$$\|\tau_2\|_{\min_c}^2 - \|\tau_1\| \|\tau_2\| \cos \theta_{\tau_1 \tau_2} \cong f_s + \nabla_{eq},$$

where  $\nabla_{eq}$  denotes a symmetric equalizer statistic and  $f_s$  denotes a centrally located statistical fulcrum.

Section 17 will examine how the total allowed eigenenergies of  $\tau_1$  and  $\tau_2$  are balanced by means of a symmetric equalizer statistic  $\nabla_{eq}$

$$\left( \|\tau_1\|_{\min_c}^2 - \|\tau_1\| \|\tau_2\| \cos \theta_{\tau_1 \tau_2} \right) + \nabla_{eq} \Leftrightarrow \left( \|\tau_2\|_{\min_c}^2 - \|\tau_1\| \|\tau_2\| \cos \theta_{\tau_1 \tau_2} \right) - \nabla_{eq}, \quad (34)$$

in relation to a centrally located statistical fulcrum  $f_s$ . Section 17 will develop statistical expressions for the symmetric equalizer statistic  $\nabla_{eq}$  and the statistical fulcrum  $f_s$ . Section 17 will also develop the statistical machinery behind a strong dual normal eigenlocus equilibrium point. The KKT complementary conditions on a strong dual normal eigenlocus  $\tau$  are examined next.

## 9.7 KKT Complementary Conditions

The KKT complementary conditions of optimization theory require that for all constraints that are not active (are not precisely met as equalities, i.e.,

$y_i (\mathbf{x}_i^T \tau + \tau_0) - 1 + \xi_i > 0$ ), the corresponding magnitudes  $\psi_i$  of the Wolfe dual normal eigenaxis components  $\psi_i \vec{\mathbf{e}}_i$  must be 0:  $\psi_i = 0$  Cristianini and Shawe-Taylor [2000], Scholkopf and Smola [2002]. It follows that Eqs (19) and (21) must be satisfied as equalities. Accordingly, let there be  $l$  Wolfe dual normal eigenaxis components  $\psi_{i*} \vec{\mathbf{e}}_i$  that have non-zero magnitudes  $\{\psi_{i*} \vec{\mathbf{e}}_i | \psi_{i*} > 0\}_{i=1}^l$  for all constraints that are precisely met as equalities.

The next section will consider the manner in which Eq. (19) determines the primal normal eigenlocus within the Wolf dual eigenspace. Equation (19) will be used to derive an expression for the  $\tau_0$  term in Eq. (22). The expression for  $\tau_0$  will then be used to obtain a normal eigenlocus test statistic for classifying unknown pattern vectors.

## 9.8 Statistical Functionality of the $\tau_0$ Term

Given Eq. (19), the following set of constrained primal normal eigenlocus equations must be satisfied as strict equalities:

$$y_i (\mathbf{x}_{i*}^T \tau + \tau_0) - 1 + \xi_i = 0, \quad i = 1, \dots, l,$$

so that an estimate for  $\tau_0$  satisfies the strong dual normal eigenlocus equation:

$$\begin{aligned} \tau_0 &= \frac{1}{l} \sum_{i=1}^l y_i (1 - \xi_i) - \frac{1}{l} \sum_{i=1}^l \mathbf{x}_{i*}^T \tau, \\ &= \frac{1}{l} \sum_{i=1}^l y_i (1 - \xi_i) - \left( \frac{1}{l} \sum_{i=1}^l \mathbf{x}_{i*} \right)^T \tau, \end{aligned} \quad (35)$$

where the expression for  $\tau_0$  is comprised of class training labels and inner product statistics between the extreme training points and  $\tau$ . Section 17 will show that the  $\tau_0$  term plays a large role in balancing the total allowed eigenenergies of  $\tau_1$  and  $\tau_2$ ; Section 17 will examine how  $\tau_0$  determines the symmetric equalizer statistic  $\nabla_{eq}$  in Eq. (34).

The expression for  $\tau_0$  in Eq. (35) is now used to obtain a normal eigenlocus test statistic that is used to classify unknown pattern vectors.

## 9.9 The Normal Eigenlocus Test Statistic

Substitution of the expression for  $\tau_0$  in Eq. (35) into the expression for the discriminant function  $D(\mathbf{x})$  in Eq. (22) provides a normal eigenlocus test statistic

$\Lambda_\tau(\mathbf{x}) \stackrel{H_1}{\underset{H_2}{\geq}} 0$  for classifying an unknown pattern vector  $\mathbf{x}$ :

$$\begin{aligned} \Lambda_\tau(\mathbf{x}) &= \mathbf{x}^T \tau - \frac{1}{l} \sum_{i=1}^l \mathbf{x}_{i*}^T \tau \\ &\quad + \frac{1}{l} \sum_{i=1}^l y_i (1 - \xi_i), \\ &= (\mathbf{x} - \bar{\mathbf{x}}_{i*})^T \tau \\ &\quad + \frac{1}{l} \sum_{i=1}^l y_i (1 - \xi_i), \end{aligned} \quad (36)$$

where the statistic  $\bar{\mathbf{x}}_{i*}$  determines the expected locus of a set of extreme training points and the statistic  $\frac{1}{l} \sum_{i=1}^l y_i (1 - \xi_i)$  accounts for the class membership of the normal eigenaxis components on  $\tau_1$  and  $\tau_2$ . An expression is now obtained for a statistical decision locus which specifies the likelihood that an unknown pattern vector belongs to a pattern category. The expression provides geometric insight into the statistical machinery of the normal eigenlocus discriminant function.

### 9.9.1 Statistical Decision Locus

Denote the unit normal eigenlocus  $\tau / \|\tau\|$  by  $\hat{\tau}$ . Letting  $\tau = \tau / \|\tau\|$  in Eq. (36) provides an expression for a statistical decision locus

$$\Lambda_{\hat{\tau}}(\mathbf{x}) = (\mathbf{x} - \bar{\mathbf{x}}_{i*})^T \tau / \|\tau\| + \frac{1}{l \|\tau\|} \sum_{i=1}^l y_i (1 - \xi_i),$$

which is determined by the scalar projection of  $\mathbf{x} - \bar{\mathbf{x}}_{i*}$  onto  $\hat{\tau}$ . More specifically, the component of  $\mathbf{x} - \bar{\mathbf{x}}_{i*}$  along  $\hat{\tau}$  determines a signed magnitude  $\|\mathbf{x} - \bar{\mathbf{x}}_{i*}\| \cos \theta$  along the axis of  $\hat{\tau}$ , where  $\theta$  is the angle between the vector  $\mathbf{x} - \bar{\mathbf{x}}_{i*}$  and  $\hat{\tau}$ . Accordingly, the component  $\text{comp}_{\hat{\tau}} \left( \overrightarrow{(\mathbf{x} - \bar{\mathbf{x}}_{i*})} \right)$  of the vector transform  $\mathbf{x} - \bar{\mathbf{x}}_{i*}$  of an unknown pattern vector  $\mathbf{x}$  along the axis of the unit normal eigenlocus  $\hat{\tau}$

$$\text{comp}_{\hat{\tau}} \left( \overrightarrow{(\mathbf{x} - \bar{\mathbf{x}}_{i*})} \right) = \|\mathbf{x} - \bar{\mathbf{x}}_{i*}\| \cos \theta,$$

determines a statistical locus  $P_{D(\mathbf{x})}$  of a category decision, where  $P_{D(\mathbf{x})}$  is at a distance of  $\|\mathbf{x} - \bar{\mathbf{x}}_{i*}\| \cos \theta$  from the origin, along the axis of the strong dual normal eigenlocus  $\tau$ . Given Eqs (23), (24), and (25), it is concluded that the scaled  $1 / \|\tau\|$  discriminant function  $\Lambda_{\hat{\tau}}(\mathbf{x})$  generates a statistical decision locus  $P_{D(\mathbf{x})}$  which lies in one of the decision regions delineated by the constrained discriminant function  $D(\mathbf{x})$  in Eq. (22).

The expression for a statistical decision locus  $P_{D(\mathbf{x})}$  provides insight into how the discriminant function  $D(\mathbf{x})$  in Eq. (22) assigns an unknown pattern vector to a pattern class. Given Eqs (23), (24), and (25), it follows that the statistic  $\text{comp}_{\hat{\tau}} \left( \overrightarrow{(\mathbf{x} - \bar{\mathbf{x}}_{i*})} \right)$  delineates a statistical decision locus  $P_{D(\mathbf{x})}$  which lies in a geometric region that is either (1) inside one of the symmetrical decision regions of large covariance depicted in Figs 17, 19, and 20, (2) on the other side of the linear decision border  $D_1(\mathbf{x})$ , where  $\tau^T \mathbf{x} + \tau_0 = +1$ , or (3) on the other side of the linear decision border  $D_{-1}(\mathbf{x})$ , where  $\tau^T \mathbf{x} + \tau_0 = -1$ . It is concluded that the statistic  $\text{comp}_{\hat{\tau}} \left( \overrightarrow{(\mathbf{x} - \bar{\mathbf{x}}_{i*})} \right)$  generates a statistical decision locus  $P_{D(\mathbf{x})}$  which specifies a likelihood that an unknown pattern vector belongs to category one or category two.

Again, letting  $\tau = \tau / \|\tau\|$  in Eq. (36), the scaled  $1 / \|\tau\|$  discriminant function  $\Lambda_{\hat{\tau}}(\mathbf{x})$

$$\Lambda_{\hat{\tau}}(\mathbf{x}) = \text{comp}_{\hat{\tau}} \left( \overrightarrow{(\mathbf{x} - \bar{\mathbf{x}}_{i*})} \right) + \frac{1}{l \|\tau\|} \sum_{i=1}^l y_i (1 - \xi_i),$$

generates an output based on the statistic  $\text{comp}_{\hat{\tau}}(\mathbf{x} - \bar{\mathbf{x}}_{i*})$  and the class membership statistic  $\frac{1}{l\|\tau\|} \sum_{i=1}^l y_i (1 - \xi_i)$ . It follows that the likelihood that an unknown pattern vector  $\mathbf{x}$  belongs to a pattern class is a function of a statistical locus  $\text{comp}_{\hat{\tau}}(\mathbf{x} - \bar{\mathbf{x}}_{i*})$  of a category decision and a class membership statistic  $\frac{1}{l\|\tau\|} \sum_{i=1}^l y_i (1 - \xi_i)$ . Figure 21 depicts a statistical decision locus generated by the discriminant function  $\Lambda_{\hat{\tau}}(\mathbf{x})$ .

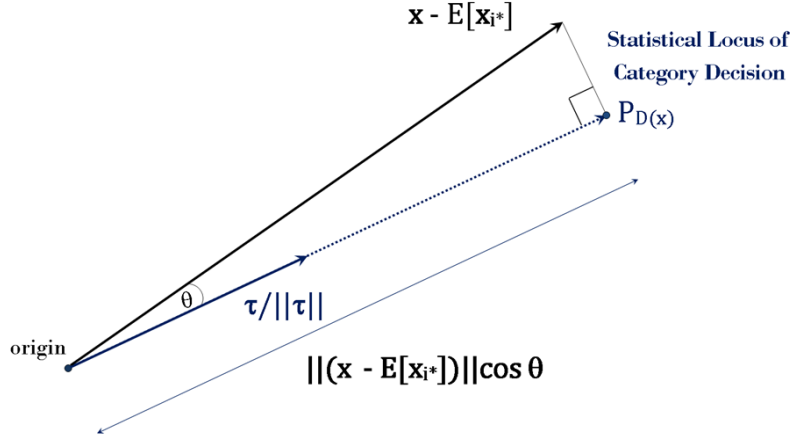


Figure 21: Illustration of a statistical decision locus  $P_{D(\mathbf{x})}$  generated by a scaled  $1/\|\tau\|$  discriminant function  $\Lambda_{\hat{\tau}}(\mathbf{x})$  for an unknown transformed pattern vector  $\mathbf{x} - E[\mathbf{x}_{i*}]$  that is projected onto  $\tau/\|\tau\|$ . Values of the statistical decision locus  $\text{comp}_{\hat{\tau}}(\mathbf{x} - \bar{\mathbf{x}}_{i*})$  and the class membership statistic  $\frac{1}{l\|\tau\|} \sum_{i=1}^l y_i (1 - \xi_i)$  specify the likelihood that the unknown vector  $\mathbf{x}$  belongs to class one or class two.

Using the expression for the discriminant function in Eq. (36), the strong dual statistical decision function  $\text{sign}(\Lambda_{\tau}(\mathbf{x}))$

$$\text{sign}(\Lambda_{\tau}(\mathbf{x})) = \text{sign} \left[ \left( \mathbf{x} - \frac{1}{l} \sum_{i=1}^l \mathbf{x}_{i*} \right)^T \tau + \dots \right] \\ \text{sign} \left[ \dots + \frac{1}{l} \sum_{i=1}^l y_i (1 - \xi_i) \right],$$

where  $\text{sign}(x) \equiv \frac{x}{|x|}$  for  $x \neq 0$ , classifies an unknown pattern vector  $\mathbf{x}_{1_i}$  or  $\mathbf{x}_{2_i}$  into category one if  $\text{sign}(\Lambda_{\tau}(\mathbf{x})) = 1$  and into category two if  $\text{sign}(\Lambda_{\tau}(\mathbf{x})) = -1$ .

Substitution of the expression for  $\tau$  of Eq. (30) into the normal eigenlocus test statistic in Eq. (36) provides a normal eigenlocus test statistic in terms of the strong dual normal eigenlocus components  $\tau_1$  and  $\tau_2$ :

$$\begin{aligned}\Lambda_{\tau_1-\tau_2}(\mathbf{x}) &= \left( \mathbf{x} - \frac{1}{l} \sum_{i=1}^l \mathbf{x}_{i*} \right)^T \tau_1 \\ &\quad - \left( \mathbf{x} - \frac{1}{l} \sum_{i=1}^l \mathbf{x}_{i*} \right)^T \tau_2 \\ &\quad + \frac{1}{l} \sum_{i=1}^l y_i (1 - \xi_i) \underset{H_2}{\overset{H_1}{\geq}} 0.\end{aligned}$$

Section 18 will examine the robust likelihood ratio that is encoded within the normal eigenlocus test statistic  $\Lambda_{\tau_1-\tau_2}(\mathbf{x})$ .

Strong dual normal eigenlocus transforms generate regularized, data-driven geometric architectures that encode robust decision statistics for any two data distributions. It will now be demonstrated that strong dual normal eigenlocus transforms determine unforeseen optimal decision systems for completely overlapping data distributions.

### 9.10 Strong Dual Normal Eigenlocus Transforms for Homogeneous Distributions

Take two Gaussian data sets that are characterized by identical means and covariance matrices. Let both pattern classes have the covariance matrix:

$$\Sigma_1 = \Sigma_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},$$

and the mean vector

$$\mu_1 = \mu_2 = (0, \ 0)^T,$$

where the Bayes' discriminant function has an error rate of 50%. Before the strong dual decision system for the homogeneous data sets outlined above is revealed, a few remarks are in order.

Recall that the constrained Lagrangian functional  $L_{\Psi(\tau)}$  of Eq. (15) returns the minimum number of normal eigenaxis components that are necessary to symmetrically partition a two class feature space. By definition, an extreme data point is located somewhere between a pair of data distributions. Given that the above data distributions are completely overlapping, it follows that all of the training vectors are extreme data points. Therefore, almost identical sets of eigen-scales will be determined for each pattern class, resulting in similar constrained primal normal eigenaxis components on  $\tau_1$  and  $\tau_2$ . Given that  $\tau_1$  and  $\tau_2$  are formed by similar normal eigenaxis components, it follows that  $\tau_1 \approx \tau_2$ . A strong dual decision system was obtained for the homogeneous data sets outlined above. The results are summarized below.

300 training vectors were obtained for each identical data category. The complete data set of 600 training vectors were transformed into a strong dual

normal eigenlocus of constrained primal normal eigenaxis components by solving the inequality constrained optimization problem of Eq. (13). As expected, all 600 training vectors were transformed into constrained primal normal eigenaxis components on  $\tau_1$  and  $\tau_2$ :

$$\begin{aligned}\tau &= \tau_1 - \tau_2, \\ &= \sum_{i=1}^{300} \psi_{1i*} \mathbf{x}_{1i*} - \sum_{i=1}^{300} \psi_{2i*} \mathbf{x}_{2i*},\end{aligned}$$

where  $\tau_1 \approx \tau_2$ . Given that  $\tau_1 \approx \tau_2$ , the width  $\frac{1}{\|\tau_1 - \tau_2\|}$  of the geometric regions between the linear decision boundary and the linear decision borders is extremely large; the distance  $\frac{2}{\|\tau_1 - \tau_2\|}$  between the linear decision borders is also extremely large. Figure 22 illustrates that strong dual decision systems determine symmetrical linear partitions of completely overlapping data distributions, where each extreme data point is enclosed in a blue circle.

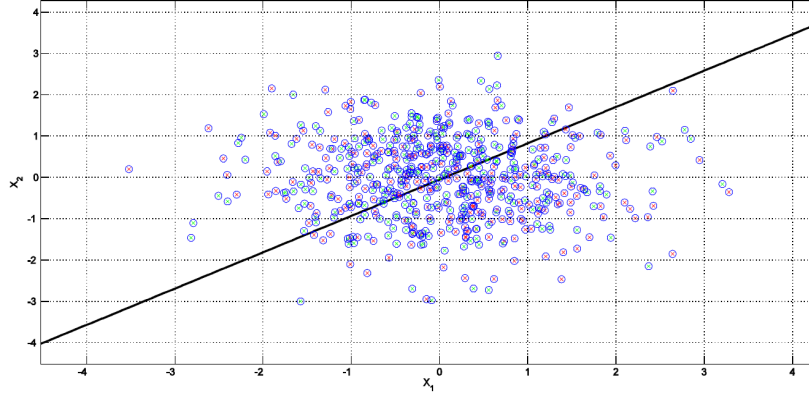


FIGURE 22. Illustration that strong dual normal eigenlocus transforms determine symmetrical linear partitions of homogeneous data distributions which are completely overlapping with each other.

Because the breadth of the geometric regions between the linear decision boundary and its borders is too large to be depicted, only the linear decision boundary can be seen in Fig. 22. The strong dual statistical decision system depicted in Fig. 22 achieves the Bayes error rate of 50%.

### 9.11 Equilibrium States of Strong Dual Decision Systems

The next two sections will outline two interrelated KKT constraints that will play a large role in determining the regularized geometric configurations of  $\psi$  and  $\tau$ . Later on, Sections 14 and 15 will examine how the KKT constraint of Eq. (38) determines a state of statistical equilibrium in which the normal eigenaxis

components on  $\psi$  and  $\tau$  are jointly and symmetrically distributed over the constrained primal normal eigenaxis components on  $\tau$ . Section 17 will demonstrate that, given any strong dual decision system in a state of statistical equilibrium, the magnitudes  $\psi_{i*}$  of the Wolfe dual normal eigenaxis components  $\psi_{i*} \vec{e}_i$  determine symmetrical eigen-scales, for which joint eigenenergies of  $\psi$  and  $\tau$  are symmetrically distributed over the eigen-scaled extreme points on  $\tau = \tau_1 - \tau_2$ , such that the total allowed eigenenergies of  $\tau_1$  and  $\tau_2$  are symmetrically balanced with each other. Therefore, given a strong dual decision system in a state of statistical equilibrium, all of the normal eigenaxis components on  $\psi$  and  $\tau$  exhibit critical lengths which satisfy the state of statistical equilibrium. Section 17 will develop and use all of these results to define the statistical equilibrium point of a strong dual decision system.

### 9.12 Critical Minimum Eigenenergy Constraints on $\tau_1$ , $\tau_2$ , and $\tau$

Returning to the KKT constraint of Eq. (21), the following algebraic system of constrained primal normal eigenlocus equations must be satisfied as strict equalities:

$$\psi_{i*} \{y_i (\mathbf{x}_{i*}^T \tau + \tau_0) - 1 + \xi_i\} = 0, \quad i = 1, \dots, l. \quad (37)$$

Section 17 will examine how the algebraic system of  $l$  strong dual normal eigenlocus equations in Eq. (37) determine critical minimum eigenenergy constraints that are satisfied by the constrained primal normal eigenlocus components  $\tau_1 - \tau_2$  on  $\tau$ . Equation (37) will be used to develop expressions for the symmetric equalizer statistic  $\nabla_{eq}$  and the implicit statistical fulcrum  $f$  in Eq. (34). The KKT constraint of Eq. (17) is examined next.

### 9.13 Equilibrium Constraints on Wolfe Dual Normal Eigenaxis Components

The KKT constraint of Eq. (17) specifies that the magnitudes of all of the Wolfe dual normal eigenaxis components on  $\psi$  must satisfy the strong dual normal eigenlocus equation:

$$(y_i = 1) \sum_{i=1}^{l_1} \psi_{1i*} + (y_i = -1) \sum_{i=1}^{l_2} \psi_{2i*} = 0,$$

so that the integrated lengths of the Wolfe dual normal eigenaxes correlated with each pattern category:

$$\sum_{i=1}^{l_1} \psi_{1i*} - \sum_{i=1}^{l_2} \psi_{2i*} = 0,$$

must balance each other:

$$\sum_{i=1}^{l_1} \psi_{1i*} = \sum_{i=1}^{l_2} \psi_{2i*}. \quad (38)$$



Section 17 will demonstrate that Eq. (38) determines a state of statistical equilibrium, for which correlated normal eigenaxis components on  $\psi$  and  $\tau$  possess critical magnitudes or lengths. It will be shown that the lengths of the Wolfe dual normal eigenaxis components must be selected so that the total allowed eigenenergies of  $\tau_1$  and  $\tau_2$  are balanced by means of a symmetric equalizer statistic  $\nabla_{eq}$  in relation to a centrally located statistical fulcrum  $f_s$ . Section 18 will demonstrate how this statistical balancing feat enables the constrained discriminant function  $D(\mathbf{x}) = \mathbf{x}^T \tau + \tau_0$  of Eq. (22) to delineate centrally located, bipartite, congruent geometric regions of large covariance for a wide variety of data distributions. Sections 14, 15, and 16 will demonstrate how the KKT constraint of Eq. (38) enforces joint symmetrical distributions of the components of  $\psi$  and  $\tau$  over each of the  $l$  constrained primal normal eigenaxis components on  $\tau$ , whereby Section 17 will demonstrate that each constrained primal normal eigenaxis component on  $\tau$  encodes an eigen-balanced eigenlocus of an extreme data point, such that the total allowed eigenenergies of  $\tau_1$  and  $\tau_2$  are symmetrically balanced with each other.

The next stage of the analysis will examine the strong dual normal eigenlocus problem within the context of the eigenlocus equation of a Wolfe dual normal eigenlocus. Section 10 will define the eigenlocus equation of a Wolfe dual normal eigenlocus. Section 10 will also examine the geometric essence and the fundamental properties of a Wolfe dual normal eigenlocus.

## 10 Eigenlocus Equation of a Wolfe Dual Normal Eigenlocus

This stage of the analysis returns to the six KKT constraints on the Lagrangian functional  $L_{\Psi(\tau)}$  of Eq. (15) which are specified by Eqs (16), (17), (18), (19), (20), and (21). The resulting expressions for a primal normal eigenlocus  $\tau$  and regularization parameters  $\xi_i$  and  $C$ , in terms of a Wolfe dual normal eigenlocus  $\psi$ , are substituted into the Lagrangian functional  $L_{\Psi(\tau)}$  of Eq. (15) and simplified. This produces the eigenlocus equation of a Wolfe dual normal eigenlocus:

$$\max \Xi(\psi) = \sum_{i=1}^N \psi_i - \sum_{i,j=1}^N \psi_i \psi_j y_i y_j \frac{[\mathbf{x}_i^T \mathbf{x}_j + \delta_{ij}/C]}{2}, \quad (39)$$

which is subject to the algebraic constraints that  $\sum_{i=1}^N y_i \psi_i = 0$  and  $\psi_i \geq 0$ , where  $\delta_{ij}$  is the Kronecker  $\delta$  defined as unity for  $i = j$  and 0 otherwise.

Equation (39) can be written in vector notation by letting  $\mathbf{Q} \triangleq \epsilon \mathbf{I} + \tilde{\mathbf{X}} \tilde{\mathbf{X}}^T$  and  $\tilde{\mathbf{X}} \triangleq \mathbf{D}_y \mathbf{X}$ , where  $\mathbf{D}_y$  is an  $N \times N$  diagonal matrix of training labels  $y_i$  and the  $N \times d$  data matrix is  $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N)^T$ . This produces the matrix version of an equation of a primal normal eigenlocus in a Wolfe dual eigenspace:

$$\max \Xi(\psi) = \mathbf{1}^T \psi - \frac{\psi^T \mathbf{Q} \psi}{2}, \quad (40)$$

which is subject to the algebraic constraints  $\psi^T \mathbf{y} = 0$  and  $\psi_i \geq 0$  Reeves [2009]. It will be assumed that the  $N$ -dimensional vector  $\psi$  whose components  $\psi_{i*} \vec{\mathbf{e}}_i$  satisfy Eqs (39) and (40) is the Wolfe dual normal eigenlocus of a hyperplane decision surface in  $\mathbb{R}^N$  that is bounded by bilaterally symmetrical hyperplane borders. Sections 14 and 15 will consider how symmetrical Wolfe dual normal eigenaxis components  $\psi_{i*} \vec{\mathbf{e}}_i$  on a Wolfe dual normal eigenlocus  $\psi$  determine the locus of a separating hyperplane  $D_{h_0}(\mathbf{x})$  that is bounded by a pair of bilaterally symmetrical hyperplane borders  $D_{h_1}(\mathbf{x})$  and  $D_{h_{-1}}(\mathbf{x})$ . Section 11 will examine how the geometric configurations of  $D_{h_0}(\mathbf{x})$ ,  $D_{h_1}(\mathbf{x})$ , and  $D_{h_{-1}}(\mathbf{x})$  are determined by the eigenspectrum of  $\mathbf{Q}$ . It will shortly be demonstrated how the constraint  $\psi^T \mathbf{y} = 0$  effectively determines the eigenlocus of  $\psi$ .

Now consider any Wolfe dual normal eigenaxis component  $\psi_{i*} \vec{\mathbf{e}}_i$  on  $\psi$ , where  $\psi_{i*} > 0$ . It will be assumed that each Wolfe dual normal eigenaxis component  $\psi_{i*} \vec{\mathbf{e}}_i$  on  $\psi$  is correlated with a  $d$ -dimensional extreme training vector  $\mathbf{x}_{i*}$ , which determines the direction of a constrained primal normal eigenaxis component  $\psi_{i*} \mathbf{x}_{i*}$  on  $\tau$ . Later on, Sections 14 and 15 will examine uniform geometric and statistical properties which are jointly exhibited by the Wolfe dual normal eigenaxis components  $\psi_{i*} \vec{\mathbf{e}}_i$  on  $\psi$  and the constrained primal normal eigenaxis components  $\psi_{i*} \mathbf{x}_{i*}$  on  $\tau$ . Sections 14 and 15 will demonstrate how the length  $\psi_{i*}$  of each Wolfe dual normal eigenaxis component  $\psi_{i*} \vec{\mathbf{e}}_i$  on  $\psi$  determines the length  $\psi_{i*} \|\mathbf{x}_{i*}\|$  of a correlated, constrained primal normal eigenaxis component  $\psi_{i*} \mathbf{x}_{i*}$  on  $\tau$ . Sections 14 and 15 will also demonstrate that the direction of each Wolfe dual normal eigenaxis component  $\psi_{i*} \vec{\mathbf{e}}_i$  is identical to the direction of a correlated, constrained primal normal eigenaxis component  $\psi_{i*} \mathbf{x}_{i*}$ . Thereby, the eigenloci of the Wolfe dual normal eigenaxis components  $\psi_{1i*} \vec{\mathbf{e}}_i$  and  $\psi_{2i*} \vec{\mathbf{e}}_i$  will be shown to determine well-proportioned eigen-scales  $\psi_{1i*}$  and  $\psi_{2i*}$  for the constrained primal normal eigenaxis components  $\psi_{1i*} \mathbf{x}_{1i*}$  and  $\psi_{2i*} \mathbf{x}_{2i*}$  on the strong dual normal eigenlocus components  $\tau_1$  and  $\tau_2$  respectively. It will be shown that each eigen-scaled extreme training point  $\psi_{1i*} \mathbf{x}_{1i*}$  or  $\psi_{2i*} \mathbf{x}_{2i*}$  specifies an eigen-balanced geometric location of a constrained primal normal eigenaxis component on  $\tau$ .

Thus far, this paper has demonstrated that strong dual normal eigenlocus transforms generate robust statistical decision systems for a wide variety of data distributions, including completely overlapping distributions. Section 17 will show that the regularized, data-driven geometric architecture depicted in Fig. 20 is configured by enforcing joint symmetrical distributions of the eigenenergies of  $\psi$  and  $\tau$  over the eigen-scaled extreme training vectors on  $\tau_1$  and  $\tau_2$ , whereby the eigenenergies of the strong dual normal eigenlocus components  $\tau_1$  and  $\tau_2$  on  $\tau$  are symmetrically balanced with each other.

The chain of arguments outlined above will be used to demonstrate how integrated, eigen-balanced sets of constrained primal normal eigenaxis components on a strong dual normal eigenlocus  $\tau = \tau_1 - \tau_2$  provide an estimate of the real unknowns, which are the constrained eigen-coordinate locations of an unknown normal eigenaxis  $\mathbf{v}$  that provides an axis of symmetry for a statistical decision system of linear partitions.

## 10.1 The Wolfe Dual Normal Eigenlocus

The geometric and statistical properties exhibited by a Wolfe dual normal eigenlocus are examined next. These properties are specified by Eqs (39) and (40), strong duality relationships between the algebraic systems of  $\max \Xi(\psi)$  and  $\min \Psi(\tau)$ , and the KKT constraints on the Lagrangian functional  $L_{\Psi(\tau)}$ . The first property concerns the geometric nature of the second-degree homogeneous polynomial surface in Eq. (39).

### 10.1.1 Second-degree Homogeneous Polynomial Surfaces

The Wolfe dual normal eigenlocus  $\psi$  of Eq. (39) is determined by a constrained polynomial equation of the form

$$\sum_{i,j=1}^N q_{ij} x_i x_j, \quad (41)$$

which is a second-degree homogeneous polynomial in  $N$  variables  $(x_1, x_2, \dots, x_N)^T$ . Second-degree polynomials written in vector notation  $\mathbf{x}^T \mathbf{Q} \mathbf{x}$  are commonly known as quadratic forms. Quadratic forms describe six classes of second-order surfaces that include  $N$ -dimensional circles, ellipses, hyperbolae, parabolas, lines, and points Hewson [2009]. Second-order surfaces are also known as quadratic or quadric surfaces.

It has previously been argued that the constrained quadratic form  $\psi^T \mathbf{Q} \psi$  denoted in Eq. (40) describes three, symmetrically positioned  $N$ -dimensional hyperplane partitioning surfaces, where the distance from each hyperplane surface to the origin is determined by a correlated constraint on the discriminant function of Eq. (22). Rayleigh's principle is now used to precisely define the geometric essence of  $\psi$ . Rayleigh's principle can be found in Strang [1986].

### 10.1.2 Geometric Essence of $\psi$

Rayleigh's principle guarantees that the quadratic ratio

$$r(\mathbf{Q}, \mathbf{x}) = \frac{\mathbf{x}^T \mathbf{Q} \mathbf{x}}{\mathbf{x}^T \mathbf{x}},$$

where  $\mathbf{Q}$  is an  $N \times N$  real symmetric matrix, is minimized by the last and smallest eigenvector  $\mathbf{x}_N$ , with its minimal value equal to the smallest eigenvalue  $\lambda_N$ :

$$\lambda_N = \min_{0 \neq \mathbf{x} \in \mathbb{R}^N} r(\mathbf{Q}, \mathbf{x}),$$

and is maximized by the first and largest eigenvector  $\mathbf{x}_1$ , with its maximal value equal to the largest eigenvalue  $\lambda_1$ :

$$\lambda_1 = \max_{0 \neq \mathbf{x} \in \mathbb{R}^N} r(\mathbf{Q}, \mathbf{x}),$$

where the inner product term  $\mathbf{x}^T \mathbf{x}$  in the quadratic ratio  $r(\mathbf{Q}, \mathbf{x}) / \mathbf{x}^T \mathbf{x}$  evaluates to a scalar. Raleigh's principle is also used to find principal eigenvectors  $\mathbf{x}_1$  which satisfy additional constraints such as

$$a_1 x_1 + \cdots a_N x_N = c,$$

for which

$$\lambda_1 = \max_{a_1 x_1 + \cdots a_N x_N = c} r(\mathbf{Q}, \mathbf{x}). \quad (42)$$

Raleigh's principle and the theorem for convex duality jointly show that Eq. (40) provides an estimate of the largest eigenvector  $\psi$  of a Gram matrix  $\mathbf{Q}$ , where  $\psi$  is a principal eigenaxis of three, symmetrical hyperplane partitioning surfaces associated with the constrained quadratic form  $\psi^T \mathbf{Q} \psi$ , such that  $\psi$  satisfies the constraints  $\psi^T \mathbf{y} = 0$  and  $\psi_i \geq 0$ . Sections 14, 15, 16, and 17 will examine how the strong duality relationships between the algebraic systems of  $\min \Psi(\tau)$  and  $\max \Xi(\psi)$  constrain the Wolfe dual normal eigenaxis components on  $\psi$  to be suitably proportioned so that the total allowed eigenenergies of  $\tau_1$  and  $\tau_2$  are balanced in a symmetric manner. The geometric and statistical properties of  $\psi$  are now summarized.

## 10.2 Fundamental Properties of $\psi$

The fundamental properties of  $\psi$  are now defined. It will first be argued that a Wolfe dual normal eigenlocus  $\psi$  satisfies a critical minimum eigenenergy constraint that is symmetrically related to the critical minimum eigenenergy constraint on  $\tau$ .

### 10.2.1 The Critical Minimum Eigenenergy Constraint on $\psi$

Equation (14) and the theorem for convex duality indicate that  $\psi$  satisfies a critical minimum eigenenergy constraint  $\|\psi\|_{\min_c}^2$  that is symmetrically related to the critical minimum eigenenergy constraint  $\|\tau\|_{\min_c}^2$  on  $\tau$

$$\|\psi\|_{\min_c}^2 \cong \|\tau\|_{\min_c}^2.$$

Accordingly, the functional  $\mathbf{1}^T \psi - \psi^T \mathbf{Q} \psi / 2$  in Eq. (40) is maximized by a set of eigen-balanced magnitudes

$$\sum_{i=1}^{l_1} \psi_{1i*} = \sum_{i=1}^{l_2} \psi_{2i*}$$

of Wolfe dual normal eigenaxis components  $\{\psi_{1i*} \vec{\mathbf{e}}_{1i*}\}_{i=1}^{l_1}$  and  $\{\psi_{2i*} \vec{\mathbf{e}}_{2i*}\}_{i=1}^{l_2}$ , for which the quadratic form

$$\psi^T \mathbf{Q} \psi / 2,$$

reaches its smallest possible value. This indicates that the eigen-balanced magnitudes of the Wolfe dual normal eigenaxis components on  $\psi$  are constrained to have the smallest possible lengths, such that the eigen-scaled extreme training

points on  $\tau$  posses suitable locations for which the total allowed eigenenergies of  $\tau_1$  and  $\tau_2$  are symmetrically balanced. Sections 12 - 16 will demonstrate that a Wolfe dual normal eigenlocus  $\psi$  satisfies a critical minimum eigenenergy constraint:

$$\max \psi^T \mathbf{Q} \psi = \lambda_{\max \psi} \|\psi\|_{\min_c}^2,$$

which is symmetrically related to the restriction of the primal normal eigenlocus to the Wolfe dual eigenspace. Section 17 will examine how the magnitudes of the Wolfe dual and the constrained primal normal eigenaxis components on  $\psi$  and  $\tau$  are properly proportioned, such that the eigenloci of the constrained primal normal eigenaxis components on  $\tau = \tau_1 - \tau_2$  possess locations which effectively balance the eigenenergies of  $\tau_1$  and  $\tau_2$ .

It will now be demonstrated how the KKT constraint  $\psi^T \mathbf{y} = 0$  effectively determines the statistical eigenlocus of  $\psi$ .

### 10.3 Wolfe Dual Statistical Systems of Partitioning Hyperplanes

The statistical representation of the primal normal eigenlocus  $\tau$  within the Wolfe dual eigenspace involves the KKT condition of Eq. (17):

$$\sum_{i=1}^N \psi_i y_i = 0, \quad i = 1, \dots, N,$$

which constrains the Wolfe dual normal eigenlocus  $\psi$  and the vector of training labels  $\mathbf{y}$  to be orthogonal  $\psi \perp \mathbf{y}$  so that

$$\psi^T \mathbf{y} = 0.$$

The orthogonality  $\psi \perp \mathbf{y}$  of  $\psi$  and  $\mathbf{y}$  indicates that the vector of training labels  $\mathbf{y}$  provides an implicit statistical directrix which determines an intrinsic reference axis in  $\mathbb{R}^N$ . Thereby, the statistical eigenlocus of  $\psi$  is uniquely specified by the distance from the statistical directrix  $\mathbf{y}$  to the endpoint of  $\psi$ . Given the strong duality relationships between the algebraic systems of  $\min \Psi(\tau)$  and  $\max \Xi(\psi)$ , it follows that a Wolfe dual normal eigenlocus  $\psi$  is an implicit, intrinsic reference axis for a separating hyperplane  $H_0(\mathbf{x}) \in \mathbb{R}^N$  that is bounded by bilaterally symmetrical hyperplane decision borders  $H_{+1} \in \mathbb{R}^N$  and  $H_{-1}(\mathbf{x}) \in \mathbb{R}^N$ . Figure 23 depicts a high level overview of a Wolfe dual statistical system of partitioning hyperplanes which is implicitly described by the orthogonality relationship  $\psi \perp \mathbf{y}$  between a statistical directrix  $\mathbf{y}$  and a Wolfe dual normal eigenlocus  $\psi$ .

## Wolfe Dual Statistical System of Partitioning Hyperplanes

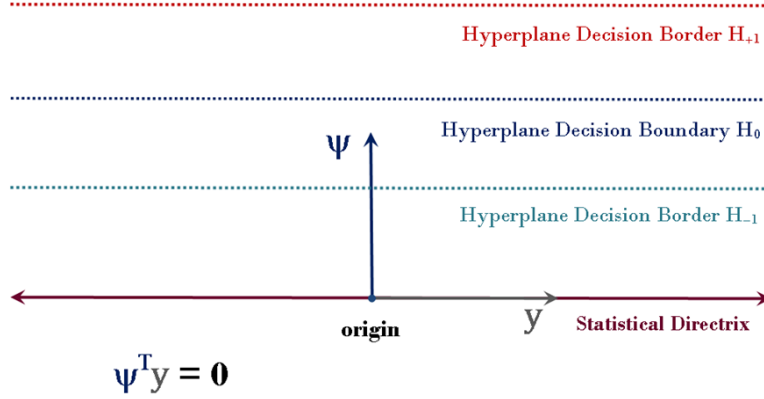


Figure 23: Illustration of a Wolfe dual statistical eigen-coordinate system of partitioning hyperplanes. All of the points on the hyperplane surfaces  $H_0$ ,  $H_{+1}$ , and  $H_{-1}$ , exclusively reference the Wolfe dual normal eigenlocus  $\psi$ , which satisfies all three hyperplane surfaces in terms of its total allowed eigenenergy.

The Wolfe dual statistical eigen-coordinate system depicted in Fig. 23 illustrates that each of the  $\psi_i$  terms returned by Eq. (40) specifies the magnitude of a normal eigenaxis component on a Wolfe dual normal eigenlocus  $\psi$ , where  $\psi$  is exclusively referenced by all three of the hyperplane surfaces specified by the Gram matrix  $\mathbf{Q}$  associated with the constrained quadratic form. The Wolfe dual normal eigenlocus  $\psi$  satisfies all three of the hyperplane surfaces in terms of its total allowed eigenenergy  $\lambda_{\max} \psi \|\psi\|_{\min_c}^2$ . Later on, Sections 14 and 15 will demonstrate that the directions of the Wolfe dual normal eigenaxis components on  $\psi$  are determined by the directions of correlated extreme training vectors.

The next section of the paper will examine how the geometric and statistical properties of strong dual normal eigenlocus transforms are sensitive to eigenspectrums of Gram matrices. Section 11 will define the principal statistical state and the characteristic eigenstates of strong dual statistical decision systems. Section 11 will consider how low rank Gram matrices cause principal statistical states and characteristic eigenstates to be substantially diminished, resulting in irregular geometric architectures which determine asymmetric linear partitions of feature spaces, resulting in ill-formed decision regions. Section 11 will also consider how the eigenspectrum of Gram matrices determines the shapes of the quadratic surfaces described by the constrained quadratic form in Eq. (40).

## 11 Weak Dual Normal Eigenlocus Transforms

This section will demonstrate how the geometric and statistical properties of strong dual normal eigenlocus transforms are sensitive to eigenspectrums of Gram matrices. It will be demonstrated that both the number and the locations of the constrained primal normal eigenaxis components on  $\tau$  are considerably affected by the rank and eigenspectrum of  $\mathbf{Q}$ . It will be shown that incomplete eigenspectrums of low rank Gram matrices  $\mathbf{Q}$  result in weak dual normal eigenlocus transforms that determine ill-formed linear decision boundaries which exhibit substandard generalization performance. It will also be shown that the geometric configurations of the dual, symmetrical linear partitioning systems in  $\mathbb{R}^N$  and  $\mathbb{R}^d$  depicted in Figs 12, 20, and 23 are largely shaped by the eigenspectrum of the Gram matrix  $\mathbf{Q}$  associated with the constrained quadratic form in Eq. (40).

### 11.1 Eigenspectrums of Gram Matrices

For pattern recognition applications where the training vector dimension  $d$  exceeds the number  $N$  of training vectors ( $d > N$ ), the solution for the Wolfe dual normal eigenlocus of Eq. (40) is well-posed, because the Gram matrix  $\mathbf{Q}$  has full rank. Machine learning solutions with eigenstructure deficiencies are generally ill-posed and ill-conditioned, and must be constrained in some manner. Numerical techniques that constrain matrix based solutions to mitigate eigenstructure deficiencies are called regularization methods Linz and Wang [2003]. For example, the Tikhonov method of regularization addresses the problem of small singular values Tikhonov and Arsenin [1977]. Regularization methods such as ridge regression and diagonal loading recondition covariance or correlation matrices Hoerl [1962].

Regularization components are essential numerical ingredients in machine learning algorithms that involve inversions of data matrices Linz [1979], Groetsch [1984], Wahba [1987], Groetsch [1993], Hansen [1998], Engl et al. [2000], Linz and Wang [2003]. The machine learning algorithm for strong dual normal eigenlocus transforms involves an inversion of the Gram matrix  $\mathbf{Q}$  in Eq. (40), so some type of regularization is required for low rank Gram matrices Reeves [2009], Reeves and Jacyna [2011].

### 11.2 Incomplete Eigenspectrums of Low Rank Gram Matrices

The solution for the Wolfe dual normal eigenlocus of Eq. (40) is ill-posed for low rank Gram matrices  $\mathbf{Q}$ , because  $\mathbf{Q}$  is singular and noninvertible. In general, learning machines that learn  $N$  parameters with  $d$  eigenfunctions have insufficient learning capacity whenever  $N > d$ . For low rank Gram matrices  $\mathbf{Q}$ , where the number  $N$  of training vectors exceeds the dimension  $d$  of the training vectors, it has been shown that a Wolfe dual normal eigenlocus  $\psi \in \mathbb{R}^N$  is spanned by an incomplete set of  $d$  eigenvectors Reeves and Jacyna [2011]. It

will now be demonstrated that low rank Gram matrices  $\mathbf{Q}$  cause the principal statistical state and the characteristic eigenstates of strong dual decision systems to be substantially diminished, resulting in irregular geometric architectures and ill-formed decision regions. It is said that diminished principal statistical states and characteristic eigenstates of strong dual decision systems produce *weak dual normal eigenlocus transforms*. Principal statistical states and characteristic eigenstates of strong dual decision systems are defined next.

### 11.3 Principal Statistical States and Characteristic Eigenstates of Strong Dual Decision Systems

Denote the principal characteristic root, i.e., the principal eigenvalue, associated with the Wolfe dual normal eigenlocus  $\psi$  of Eq. (40) by  $\lambda_{\max_\psi}$ . Define the principal characteristic root  $\lambda_{\max_\psi}$  to be the principal statistical state of a strong dual decision system. Define the characteristic eigenstates

$$\Psi_\tau(\mathbb{R}^d) = \left\{ \left\{ \Psi_{\tau_1}(\mathbb{R}^d) \right\}_{i=1}^{l_1}, \left\{ \Psi_{\tau_2}(\mathbb{R}^d) \right\}_{i=1}^{l_2} \right\},$$

of a strong dual decision system to be the eigen-scaled extreme training points on  $\tau_1$

$$\left\{ \Psi_{\tau_1}(\mathbb{R}^d) \right\}_{i=1}^{l_1} \triangleq \left\{ \psi_{1i*} \mathbf{x}_{1i*} | \mathbf{x}_{1i*} \sim p_{\mathbf{x}_{1i} | \mathbf{X}_1}(\mathbf{x}_{1i} | \mathbf{X}_1) \right\}_{i=1}^{l_1}, \quad (43)$$

and on  $\tau_2$

$$\left\{ \Psi_{\tau_2}(\mathbb{R}^d) \right\}_{i=1}^{l_2} \triangleq \left\{ \psi_{2i*} \mathbf{x}_{2i*} | \mathbf{x}_{2i*} \sim p_{\mathbf{x}_{2i} | \mathbf{X}_2}(\mathbf{x}_{2i} | \mathbf{X}_2) \right\}_{i=1}^{l_2}. \quad (44)$$

Later on, Section 18 will demonstrate that the characteristic eigenstates in Eqs (43) and (44) encode the likelihoods of finding extreme data points in particular regions of  $\mathbb{R}^d$ .

The principal statistical state  $\lambda_{\max_\psi}$  of a strong dual decision system is extensively diminished for low rank Gram matrices  $\mathbf{Q}$ . In particular, low rank Gram matrices  $\mathbf{Q}$  provide insufficient estimates of principal statistical states  $\lambda_{\max_\psi}$ , resulting in incomplete and/or defective sets of characteristic eigenstates  $\left\{ \left\{ \Psi_{\tau_1}(\mathbb{R}^d) \right\}_{i=1}^{l_1}, \left\{ \Psi_{\tau_2}(\mathbb{R}^d) \right\}_{i=1}^{l_2} \right\}$  of strong dual decision systems. Thereby, low rank Gram matrices  $\mathbf{Q}$  of linear kernel SVMs generate weak dual normal eigenlocus transforms that produce ill-formed linear decision boundaries which exhibit substandard generalization performance for overlapping data distributions. For non-overlapping data distributions, low rank Gram matrices  $\mathbf{Q}$  of linear kernel SVMs determine ill-formed linear decision boundaries which exhibit optimal generalization performance at the expense of unnecessary sets of characteristic eigenstates.

Both the number and the locations of the constrained primal normal eigenaxis components on  $\tau$  are considerably affected by the rank and eigenspectrum of  $\mathbf{Q}$ . For example, given non-overlapping data distributions and low rank Gram



matrices, *all* of the training data are transformed into normal eigenaxis components Reeves [2009]. In general, perturbations of the principal statistical states and characteristic eigenstates of strong dual decision systems produce irregular geometric architectures which determine asymmetric linear partitions of feature spaces, resulting in ill-formed decision regions. As an example, Fig. 24 depicts an asymmetric partitioning produced by a weak dual normal eigenlocus transform of training data described by the covariance matrix:

$$\Sigma_1 = \Sigma_2 = \begin{pmatrix} 7 & 0 \\ 0 & 0.5 \end{pmatrix},$$

and the mean vectors  $\mu_1 = (3, 7)^T$  and  $\mu_2 = (2, 6)^T$ , for which the linear decision boundary and its borders are badly skewed and poorly positioned.

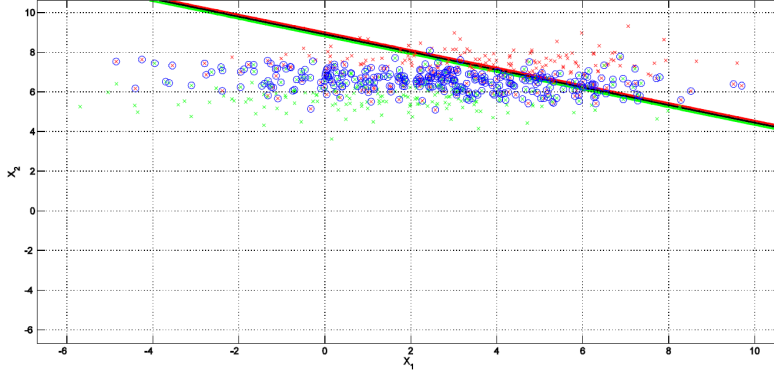


Figure 24: Illustration that weak dual normal eigenlocus transforms based on insufficient eigenstates result in asymmetric linear partitions and poorly positioned decision regions.

Only 53% of the training data are transformed into normal eigenaxis components, whereas properly regularized linear SVM transforms  $\approx 86\%$  of training data to learn this optimal partitioning Reeves [2009]. For this example, low rank Gram matrices cause asymmetrical distributions of principal eigenenergies over insufficient sets of eigen-scaled extreme data points. On the other hand, given non-overlapping data distributions and low rank Gram matrices, all of the training data are transformed into normal eigenaxis components. In both instances, low rank Gram matrices generate weak dual normal eigenlocus transforms. Additional examples of ill-formed linear decision regions resulting from weak dual normal eigenlocus transforms can be found in Reeves [2009] and Reeves and Jacyna [2011].

Given a previous analysis of  $\psi$  for low rank Gram matrices  $\mathbf{Q}$ , which can be found in Reeves and Jacyna [2011], within the context of strong dual normal eigenlocus transforms, it is concluded that incomplete sets of eigenvectors

generate incomplete eigenspectrums and insufficient eigenstates for strong dual normal eigenlocus transforms. Overall, it is concluded that Wolfe dual normal eigenlocus estimates of  $\psi$  that are based on incomplete eigenspectrums of low rank Gramian matrices  $\mathbf{Q}$  produce weak dual normal eigenlocus estimates of  $\tau$  that are based on perturbed principal statistical states and insufficient eigenstates. Generally speaking, low rank Gramian matrices  $\mathbf{Q}$  provide insufficient estimates of principal statistical states  $\lambda_{\max, \psi}$ , resulting in asymmetrical distributions of the eigenenergies of  $\psi$  and  $\tau$  over the eigen-scaled extreme data points on  $\tau$ . Figure 25 depicts the geometric and statistical connections between the joint statistical contents and the symmetrical geometric configurations of  $\psi$  and  $\tau$ .

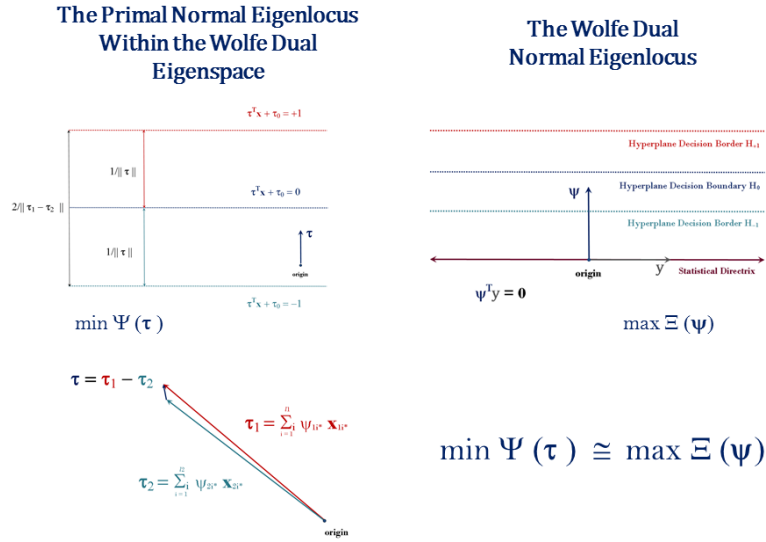


FIGURE 25. Illustration of the symmetrical geometric relationships between the constrained primal normal eigenaxis components on  $\tau$  and the Wolfe dual normal eigenaxis components on  $\psi$ . The constrained primal normal eigenlocus  $\tau = \tau_1 - \tau_2$  possesses a geometric configuration which is determined by the statistical contents of symmetrical normal eigenaxis components on its Wolfe dual  $\psi$ .

#### 11.4 Generating Sufficient Eigenspectrums for Low Rank Gram Matrices

Take any collection of  $N$  training vectors of dimension  $d$ , for which  $d < N$  and  $\mathbf{Q}$  has low rank. It has been shown that the regularized form of Eq. (40), for which  $\epsilon \ll 1$  and  $\mathbf{Q} \triangleq \epsilon \mathbf{I} + \tilde{\mathbf{X}} \tilde{\mathbf{X}}^T$ , ensures that  $\mathbf{Q}$  has full rank, and thereby ensures

that  $\mathbf{Q}$  has a complete eigenspectrum and eigenvector set. Thus, the auxiliary functional  $C/2 \sum_{i=1}^N \xi_i^2$  in Eq. (15) ensures that the matrix-based estimate of the hyperplane surfaces determined by Eq. (40) is based on a full rank Gram matrix  $\mathbf{Q}$ , so that the statistical contents of  $\psi$  are based on a complete eigenspectrum and eigenvector set. The regularization constant  $C$  in Eq. (39) is related to the regularization parameter  $\epsilon$  by  $\frac{1}{C}$  Reeves and Jacyna [2011].

Therefore, given any collection of  $N$  training vectors of dimension  $d$ , for which  $N < d$ , the Gram matrix  $\mathbf{Q}$  in Eq. (40) has full rank, and the regularization parameters  $\xi_i$  in the primal normal eigenlocus of Eq. (13) and all of its derivatives are set equal to zero:  $\xi_i = 0$ .

The eigenspectrum of  $\mathbf{Q}$  plays a fundamental role in describing the hyperplane surfaces which are implicitly delineated by  $\psi$ . The next section will demonstrate that the eigenspectrum of  $\mathbf{Q}$  determines the shapes of the quadratic surfaces described by the constrained quadratic form in Eq. (40).

## 11.5 Eigenspectrum Shaping of Quadratic Surfaces

Take the standard equation of a quadratic form:  $\mathbf{x}^T \mathbf{Q} \mathbf{x} = 1$ . Write  $\mathbf{x}$  in terms of an orthogonal basis of unit eigenvectors  $\{\mathbf{v}_1, \dots, \mathbf{v}_N\}$  so that  $\mathbf{x} = \sum_{i=1}^N x_i \mathbf{v}_i$ . Substitution of this expression into  $\mathbf{x}^T \mathbf{Q} \mathbf{x}$

$$\mathbf{x}^T \mathbf{Q} \mathbf{x} = \left( \sum_{i=1}^N x_i \mathbf{v}_i \right)^T \mathbf{Q} \left( \sum_{j=1}^N x_j \mathbf{v}_j \right)$$

produces a simple coordinate form expression of a second-order surface:

$$\lambda_1 x_1^2 + \lambda_2 x_2^2 + \dots + \lambda_N x_N^2 = 1, \quad (45)$$

solely in terms of the eigenvalues  $\lambda_N \leq \lambda_{N-1} \dots \leq \lambda_1$  of the matrix  $\mathbf{Q}$  Hewson [2009]. Equation (45) reveals that *the geometric shape of a quadratic surface is completely determined by the eigenvalues of the matrix associated with a quadratic form*. This general property of quadratic forms will lead to far reaching consequences for the strong dual normal eigenlocus method for linear decision boundary estimates and the strong dual principal eigenlocus method for second-order decision boundary estimates.

It will now be argued that the inner product statistics of a training data collection effectively determine the geometric shapes of the quadratic surfaces described by the constrained quadratic form in Eq. (40).

Consider a Gram or kernel matrix  $\mathbf{Q}$  associated with the constrained quadratic form in Eq. (40). Denote the elements of the Gram or kernel matrix  $\mathbf{Q}$  by  $\varphi(\mathbf{x}_i, \mathbf{x}_j)$ , where  $\varphi(\mathbf{x}_i, \mathbf{x}_j)$  denotes an inner product relationship between the training vectors  $\mathbf{x}_i$  and  $\mathbf{x}_j$ . The Cayley-Hamilton theorem provides the result that the eigenvalues  $\{\lambda_i\}_{i=1}^N \in \Re$  of  $\mathbf{Q}$  satisfy the characteristic equation

$$\det(\mathbf{Q} - \lambda \mathbf{I}) = 0,$$

which is a polynomial of degree  $N$ . The roots  $p(\lambda) = 0$  of the characteristic polynomial  $p(\lambda)$  of  $\mathbf{Q}$ :

$$\det \left( \begin{bmatrix} \varphi(\mathbf{x}_1, \mathbf{x}_1) - \lambda_1 & \cdots & \varphi(\mathbf{x}_1, \mathbf{x}_N) \\ \varphi(\mathbf{x}_2, \mathbf{x}_1) & \cdots & \varphi(\mathbf{x}_2, \mathbf{x}_N) \\ \vdots & \ddots & \vdots \\ \varphi(\mathbf{x}_N, \mathbf{x}_1) & \cdots & \varphi(\mathbf{x}_N, \mathbf{x}_N) - \lambda_N \end{bmatrix} \right) = 0,$$

are also the eigenvalues  $\lambda_N \leq \lambda_{N-1} \leq \dots \leq \lambda_1$  of  $\mathbf{Q}$  Lathi [1998]. Therefore, given that (1) the roots of a characteristic polynomial  $p(\lambda)$  vary continuously with its coefficients, and that (2) the coefficients of  $p(\lambda)$  can be expressed in terms of sums of principal minors Meyer [2000], it follows that the coefficients of  $p(\lambda)$ , and therefore the eigenvalues of  $\mathbf{Q}$ , vary continuously with the inner product elements  $\varphi(\mathbf{x}_i, \mathbf{x}_j)$  of  $\mathbf{Q}$ . It is concluded that the eigenvalues  $\lambda_N \leq \lambda_{N-1} \leq \dots \leq \lambda_1$  of a Gram or kernel matrix  $\mathbf{Q}$  are actually determined by its inner product elements  $\varphi(\mathbf{x}_i, \mathbf{x}_j)$ .

Given Eq. (45) and the continuous functional relationship between the inner product elements and the eigenvalues of a Gram or kernel matrix, it follows that the geometric shapes of the three, symmetrical quadratic partitioning surfaces described by Eqs (39) or (40) are an inherent function of inner product statistics  $\varphi(\mathbf{x}_i, \mathbf{x}_j)$  between training vectors.

It is concluded that the algebraic form of the inner product statistics encoded within Gram or kernel matrices effectively determines the shapes of the three, symmetrical quadratic partitioning surfaces described by Eqs (39) or (40). For strong dual normal eigenlocus transforms, given that coordinate form expressions of hyperplane surfaces involve first-degree vector components  $x_i$ , it is claimed that the algebraic form of an inner product statistic must encode first-degree vector components for effective descriptions of hyperplane surfaces.

Alternatively, given that coordinate form expressions for nonlinear second-order surfaces involve first  $x_i$  and second-degree, i.e.,  $x_i^2$  or  $x_i x_j$ , vector components, it is claimed that the algebraic form of an inner product statistic must encode both first and second-degree vector components for effective descriptions of quadratic surfaces.

Given the chain of arguments outlined above, it is concluded that the algebraic form of the inner product statistics encoded within a Gram or kernel matrix determine the geometric shapes of the three, symmetrical quadratic partitioning surfaces described by the constrained quadratic form in Eq. (40). It follows that, given a suitable algebraic form for an inner product statistic, the eigenvalues of a Gram or kernel matrix associated with the constrained quadratic form in Eq. (40) describe either  $N$ -dimensional circles, ellipses, hyperbolae, parabolas, or lines. Section 12 will argue that a Gram matrix  $\mathbf{Q}$  associated with the constrained quadratic form in Eq. (40), whose inner product elements  $\varphi(\mathbf{x}_i, \mathbf{x}_j)$  have the algebraic form of  $\mathbf{x}_i^T \mathbf{x}_j$ , encodes descriptive statistics for three, symmetrical hyperplane partitioning surfaces.

It will now be demonstrated that kernel matrices  $\mathbf{Q}$  associated with the constrained quadratic form in Eq. (40), whose inner product elements  $\varphi(\mathbf{x}_i, \mathbf{x}_j)$

have the algebraic form of  $(\mathbf{x}_i^T \mathbf{x}_j + 1)^2$ , encode descriptive statistics for three, symmetrical,  $N$ -dimensional partitioning circles, ellipses, hyperbolae, or parabolas, which are correlated to three, symmetrical,  $d$ -dimensional partitioning circles, ellipses, hyperbolae, or parabolas. The claim is demonstrated by applying second-order polynomial kernel SVMs to two sets of Gaussian data.

Second-order polynomial kernel SVMs are first applied to the overlapping Gaussian data sets of classification example two. Figure 26 illustrates a second-order decision boundary that is determined by three, symmetrical, 2-dimensional partitioning parabolas, all of which are delineated by the constrained discriminant function of a strong dual principal eigenlocus transform. All three parabolas are positioned in symmetrical locations that delineate geometric regions of data distribution overlap. Moreover, the strong dual principal decision system achieves the Bayes' error rate of 25% for this classification problem. All of the points that lie on each 2-dimensional parabola exclusively reference a common principal eigenaxis. The principal eigenaxis estimate, which is specified by the primal and Wolfe dual eigenlocus equations of a *strong dual principal eigenlocus*, involves solving an inequality constrained optimization problem that is similar in nature to Eq. (13).

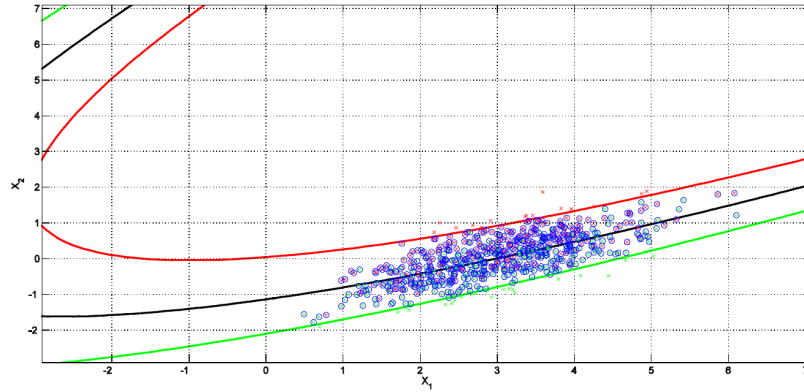


Figure 26: Illustration that a second-order polynomial kernel matrix encodes descriptive statistics for three, symmetrically positioned,  $N$ -dimensional partitioning parabolas. Thereby, polynomial kernel SVM estimates a principal eigenaxis which is exclusively referenced by all of the points on each symmetrically positioned 2-dimensional parabola, such that all three 2-dimensional parabolas jointly delineate a symmetrical partitioning of overlapping Gaussian data.

Next, second-order polynomial kernel SVMs are applied to the completely overlapping Gaussian data sets considered in Section 9. Figure 27 illustrates a second-order decision boundary that is determined by three, symmetrical, 2-dimensional partitioning hyperbolae, all of which are delineated by the constrained discriminant function of a strong dual principal eigenlocus transform.

All three hyperbolae are positioned in symmetrical locations that delineate geometric regions of data distribution overlap. All of the points that lie on each 2-dimensional hyperbola exclusively reference a common principal eigenaxis. The strong dual principal eigenlocus transform is specified by the primal and Wolfe dual eigenlocus equations of a strong dual principal eigenlocus. The strong dual principal decision system achieves the Bayes' error rate of 50% for this classification problem.

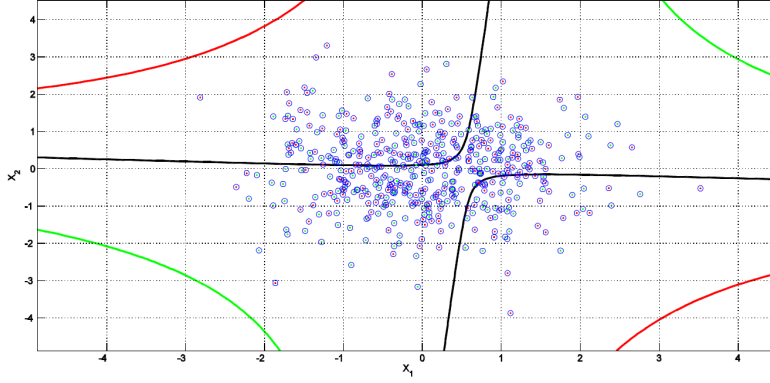


Figure 27: Illustration that a second-order polynomial kernel matrix encodes descriptive statistics for three, symmetrically positioned,  $N$ -dimensional partitioning hyperbolae. Thereby, polynomial kernel SVM estimates a principal eigenaxis, which is exclusively referenced by all of the points on each symmetrically positioned 2-dimensional hyperbola, such that all three 2-dimensional hyperbolae jointly delineate a symmetrical partitioning of completely overlapping Gaussian data.

## 11.6 Descriptive Statistics Encoded Within $\psi$

Consider the Gram matrix  $\mathbf{Q}$  associated with the constrained quadratic form in Eq. (40). The eigenvectors  $v$  of  $\mathbf{Q}$

$$\mathbf{Q}v_i = \lambda_i v_i, \quad i = 1, \dots, N,$$

correspond to directions left unchanged by the action of the Gram matrix  $\mathbf{Q}$  Meyer [2000]. This implies that the directions of the Wolfe dual normal eigenaxis components  $\psi_{i*} \vec{e}_{i*}$  on  $\psi$  are left unchanged by the inner product elements  $\mathbf{x}_i^T \mathbf{x}_j$  of  $\mathbf{Q}$ .

Suppose that  $\mathbf{Q}$  contains descriptive statistics  $\Sigma(\mathbf{x}_i, \mathbf{x}_j)$  for a hyperplane decision surface  $h_{D_0}(\mathbf{x})$  that is bounded by bilaterally symmetrical hyperplane decision borders  $h_{D_{h+1}}(\mathbf{x})$  and  $h_{D_{h-1}}(\mathbf{x})$ . Consider transforming the statistics  $\Sigma(\mathbf{x}_i, \mathbf{x}_j)$  embedded within  $\mathbf{Q}$ :

$$\mathbf{Q}\psi = \lambda_{\max \psi} \psi,$$

into a data-driven, non-orthogonal set of Wolfe dual normal eigenaxis components

$$\mathbf{Q} \sum_{i=1}^l \psi_{i*} \vec{\mathbf{e}}_{i*} = \lambda_{\max \psi} \sum_{i=1}^l \psi_{i*} \vec{\mathbf{e}}_{i*},$$

formed by  $l$  eigen-scaled  $\psi_{i*}$  non-orthogonal unit vectors  $\{\vec{\mathbf{e}}_{1*}, \dots, \vec{\mathbf{e}}_{l*}\}$ , where the eigenlocus of each Wolfe dual normal eigenaxis component  $\psi_{i*} \vec{\mathbf{e}}_{i*}$  is determined by the direction and eigen-balanced magnitude of a correlated extreme vector  $\mathbf{x}_{i*}$ .

Given the above assumptions, it will shortly be demonstrated how a Wolfe dual normal eigenlocus  $\psi$  provides an estimate of a distinctive normal eigenaxis in  $\mathbb{R}^N$  that shapes and complements the constrained primal estimate  $\tau$  of a similar normal eigenaxis in  $\mathbb{R}^d$ . An expression will be developed for a Wolfe dual normal eigenlocus  $\psi$  that contains descriptive statistics for three, symmetrical hyperplane partitioning surfaces in  $\mathbb{R}^N$ . The same expression describes point and coordinate relationships between the eigen-scaled extreme points on  $\tau$  and the Wolfe dual normal eigenaxis components on  $\psi$ . The expression will be used to identify uniform geometric and statistical properties which are jointly exhibited by correlated normal eigenaxis components on  $\psi$  and  $\tau$ .

The next section will motivate the examination of point and coordinate relationships between the constrained primal and the Wolfe dual normal eigenaxis components. Section 12 will define pointwise covariance statistics for individual training points, and will demonstrate how pointwise covariance statistics can be used to find extreme data points which possess large pointwise covariances. Section 12 will also consider the total allowed eigenenergies of a strong dual normal eigenlocus.

## 12 Point and Coordinate Relationships Between Constrained Primal and Wolfe Dual Normal Eigenaxis Components

A geometric object is assumed to be independent of the coordinate system that is used to describe it Hewson [2009]. On the contrary, this paper considers major intrinsic coordinate axes of conic sections and quadratic surfaces to be an inherent part of second-order geometric loci. An upcoming paper will substantiate the claim that the locus of a principal eigenaxis is a distinctive, invariant, and hardwired geometric property of second-order curves and surfaces, which effectively determines the points on a second-order locus.

This paper has rigorously demonstrated that the locus of a normal eigenaxis is a distinctive, invariant, and hardwired geometric property of linear curves and surfaces, which effectively determines the points on a linear locus. It has been argued that the locations of the constrained primal normal eigenaxis components on the constrained primal normal eigenlocus  $\tau$  of Eq. (13) provide estimates for the constrained eigen-coordinate locations of a normal eigenaxis  $\mathbf{v}$  of a linear decision boundary. It has been demonstrated that the constrained

normal eigenlocus  $\tau$  of Eq. (13) delineates a linear decision boundary that is bounded by bilaterally symmetrical linear decision borders. It has also been demonstrated that the statistical eigen-coordinate system of Eqs (22), (23), (24), and (25), depicted in Fig. 20, delineates bipartite, symmetric regions of large covariance located between two data distributions in  $\mathbb{R}^d$ , which describe regions of data distribution overlap for overlapping distributions and bipartite symmetric partitions between the tail regions of non-overlapping data distributions.

Thus far, this paper has argued that the scaling parameters  $\psi_{i*}$  returned by Eq. (40) determine symmetrical lengths of Wolfe dual normal eigenaxis components  $\psi_{i*} \vec{e}_{i*}$  on a Wolfe dual normal eigenlocus  $\psi$ . Additional insights can be obtained by investigating the algebraic, geometric, and statistical nature of the point and coordinate relationships between the eigen-scaled extreme points on  $\tau$  and the Wolfe dual normal eigenaxis components on  $\psi$ . In order to obtain these insights, it will be necessary to develop algebraic expressions which describe algebraic, geometric, and statistical relationships between the Wolfe dual normal eigenaxis components and the eigen-scaled extreme training points. The expressions must also describe point and coordinate relationships between the extreme training points.

Sections 13 - 15 will develop an algebraic expression for  $\psi$  that describes the point and coordinate relationships outlined above. The expression will be used to examine how each Wolfe dual normal eigenaxis component is formed by an eigen-balanced set of eigen-scaled scalar projections of extreme training vectors, along the common axis of an extreme training vector which is correlated with the Wolfe dual normal eigenaxis component. Thereby, the expression will be used to identify uniform geometrical and statistical properties which are exhibited by Wolfe dual normal eigenaxis components  $\psi_{i*} \vec{e}_i$  on  $\psi$  and correlated, constrained primal normal eigenaxis components  $\psi_{i*} \mathbf{x}_{i*}$  on  $\tau$ . It will be demonstrated that each Wolfe dual normal eigenaxis component  $\psi_{i*} \vec{e}_i$  on  $\psi \in \mathbb{R}^N$  has an eigenlocus which stores an eigen-balanced, pointwise covariance estimate  $\psi_{i*}$  of a correlated extreme data point  $\mathbf{x}_{i*} \in \mathbb{R}^d$ , such that each eigen-balanced, pointwise covariance estimate  $\psi_{i*}$  encodes an eigen-balanced first and second order statistical moment about the locus of an extreme data point  $\mathbf{x}_{i*}$ , which determines a suitable length  $\psi_{i*} \|\mathbf{x}_{i*}\|$  for a constrained primal normal eigenaxis component  $\psi_{i*} \mathbf{x}_{i*}$  on  $\tau \in \mathbb{R}^d$ .

The notion of a first and second order statistical moment about the locus of a data point will be defined next, along with the notion of a pointwise covariance estimate, both of which are shown to provide a maximum covariance estimate in a principal location.

## 12.1 Joint Statistical Underpinnings of $\psi$ and $\tau$

An algebraic expression has been obtained for  $\psi$

$$\psi = \mathbf{Q}^{-1} \left[ \mathbf{Q} - \frac{\mathbf{y}\mathbf{y}^T}{\mathbf{y}^T \mathbf{Q}^{-1} \mathbf{y}} \right] \mathbf{Q}^{-1} (\mathbf{1} + \lambda),$$



that relates the Wolfe dual normal eigenaxis components to inner product statistics between the training vectors stored within  $\mathbf{Q}$  Reeves [2009], Reeves and Jacyna [2011]. The above expression for  $\psi$  clearly illustrates that the Wolfe dual normal eigenlocus solution of Eq. (40) is ill-posed for singular and noninvertible  $\mathbf{Q}$ . The expression is a nonlinear functional of  $\mathbf{y}$ ,  $\mathbf{Q}$ , and  $\mathbf{Q}^{-1}$  that generally involves intractable point and coordinate relationships between the training data. Therefore, the above algebraic expression cannot be used to investigate the algebraic, geometric, or statistical nature of the point and coordinate relationships between the eigen-scaled extreme points and the Wolfe dual normal eigenaxis components.

However, it can be investigated how the magnitudes and the directions of the Wolfe dual normal eigenaxis components on  $\psi$  are selected to minimize the value of the quadratic form  $\psi^T \mathbf{Q} \psi$  in Eq. (40). To accomplish this, an algebraic connection will be exploited, between the quadratic form  $\psi^T \mathbf{Q} \psi$  in Eq. (40) and the critical minimum eigenenergies of  $\psi$  and  $\tau$ , where the algebraic connection involves a principal eigen-decomposition of  $\mathbf{Q}$ .

An algebraic expression for a principal eigen-decomposition of  $\mathbf{Q}$  will be developed that offers tractable point and coordinate relationships between the eigen-scaled extreme training points on  $\tau$  and the Wolfe dual normal eigenaxis components on  $\psi$ . The expression will be used to demonstrate how eigenloci of Wolfe dual normal eigenaxis components and constrained primal normal eigenaxis components are determined by eigen-balanced vector components along the axes of extreme vectors. The expression will also be used to demonstrate that Wolfe dual normal eigenaxis components on  $\psi$  and correlated constrained primal normal eigenaxis components on  $\tau$  possess symmetrical lengths and exhibit directional symmetry, which jointly describe principal locations of large covariance, whereby the constrained discriminant function  $D(\mathbf{x}) = \mathbf{x}^T \tau + \tau_0$  delineates centrally located, bipartite, symmetric regions of large covariance between two data distributions.

The next section will examine how first and second order statistical moments of data points are encoded within Gramian matrices. The section begins with distributions of first degree vector coordinates.

## 12.2 Distributions of First Degree Vector Coordinates

Consider again the Gramian matrix  $\mathbf{Q}$  associated with the constrained quadratic form in Eq. (40)

$$\mathbf{Q} = \begin{pmatrix} \mathbf{x}_1^T \mathbf{x}_1 & \mathbf{x}_1^T \mathbf{x}_2 & \cdots & -\mathbf{x}_1^T \mathbf{x}_N \\ \mathbf{x}_2^T \mathbf{x}_1 & \mathbf{x}_2^T \mathbf{x}_2 & \cdots & -\mathbf{x}_2^T \mathbf{x}_N \\ \vdots & \vdots & \ddots & \vdots \\ -\mathbf{x}_N^T \mathbf{x}_1 & -\mathbf{x}_N^T \mathbf{x}_2 & \cdots & \mathbf{x}_N^T \mathbf{x}_N \end{pmatrix}, \quad (46)$$

where  $\mathbf{Q} \triangleq \tilde{\mathbf{X}} \tilde{\mathbf{X}}^T$ ,  $\tilde{\mathbf{X}} \triangleq \mathbf{D}_y \mathbf{X}$ ,  $\mathbf{D}_y$  is a  $N \times N$  diagonal matrix of training labels  $y_i$  and the  $N \times d$  data matrix is  $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N)^T$ . Without loss of

generality, let  $N$  be an even number. Let the first  $N/2$  vectors have the training label  $y_i = 1$  and the last  $N/2$  vectors have the training label  $y_i = -1$ .

Given the above assumptions, the Gramian matrix  $\mathbf{Q}$  stores a highly structured collection of inner product statistics  $\mathbf{x}_i^T \mathbf{x}_j$  between the geometric loci of the  $N$  training points stored within  $\tilde{\mathbf{X}}$ . Take the training point  $\mathbf{x}_i$  or  $\mathbf{x}_j$ , along with the constraint that index  $i = j$ . It follows that row  $\mathbf{Q}(i, :)$  or column  $\mathbf{Q}(:, j)$  encodes sample inner product statistics between the vector  $\mathbf{x}_i$  or  $\mathbf{x}_j$  and all of the vectors  $(\mathbf{x}_1, \dots, \mathbf{x}_N)$  in a training data collection. It will now be shown that inner product statistics encoded within Gram matrices determine distributions of first degree vector coordinates. At this stage of the analysis, training labels will not be taken into account.

Take the training points  $\mathbf{x}_i$  and  $\mathbf{x}_j$ , along with the constraint that index  $i = j$ . Using the algebraic relationship

$$\mathbf{x}_i^T \mathbf{x}_j = \|\mathbf{x}_i\| \|\mathbf{x}_j\| \cos \theta_{\mathbf{x}_i \mathbf{x}_j},$$

satisfied by the inner product statistic  $\mathbf{x}_i^T \mathbf{x}_j$ , it follows that row  $\mathbf{Q}(i, :)$  in Eq. (46) encodes uniformly weighted  $\|\mathbf{x}_i\|$  scalar projections  $\|\mathbf{x}_j\| \cos \theta_{\mathbf{x}_i \mathbf{x}_j}$  for each of the  $N$  vectors  $\{\mathbf{x}_j\}_{j=1}^N$  onto the vector  $\mathbf{x}_i$ :

$$\tilde{\mathbf{Q}} = \begin{pmatrix} \|\mathbf{x}_1\| \|\mathbf{x}_1\| \cos \theta_{\mathbf{x}_1 \mathbf{x}_1} & \cdots & -\|\mathbf{x}_1\| \|\mathbf{x}_N\| \cos \theta_{\mathbf{x}_1 \mathbf{x}_N} \\ \|\mathbf{x}_2\| \|\mathbf{x}_1\| \cos \theta_{\mathbf{x}_2 \mathbf{x}_1} & \cdots & -\|\mathbf{x}_2\| \|\mathbf{x}_N\| \cos \theta_{\mathbf{x}_2 \mathbf{x}_N} \\ \vdots & \ddots & \vdots \\ -\|\mathbf{x}_N\| \|\mathbf{x}_1\| \cos \theta_{\mathbf{x}_N \mathbf{x}_1} & \cdots & \|\mathbf{x}_N\| \|\mathbf{x}_N\| \cos \theta_{\mathbf{x}_N \mathbf{x}_N} \end{pmatrix}, \quad (47)$$

where  $0 < \theta_{\mathbf{x}_i \mathbf{x}_j} \leq \frac{\pi}{2}$  or  $\frac{\pi}{2} < \theta_{\mathbf{x}_i \mathbf{x}_j} \leq \pi$ . Alternatively, column  $\mathbf{Q}(:, j)$  in Eq. (46) encodes weighted  $\|\mathbf{x}_i\|$  scalar projections  $\|\mathbf{x}_j\| \cos \theta_{\mathbf{x}_i \mathbf{x}_j}$  for the vector  $\mathbf{x}_j$  onto each of the  $N$  vectors  $\{\mathbf{x}_i\}_{i=1}^N$ .

### 12.2.1 Signed Magnitudes of Vector Projections

Now consider the  $i$ th row  $\tilde{\mathbf{Q}}(i, :)$  of  $\tilde{\mathbf{Q}}$  in Eq. (47). Given Eq. (12), it follows that element  $\tilde{\mathbf{Q}}(i, j)$  of row  $\tilde{\mathbf{Q}}(i, :)$  encodes the length of the vector  $\mathbf{x}_i$  multiplied by the scalar projection of  $\mathbf{x}_j$  onto  $\mathbf{x}_i$ :

$$\tilde{\mathbf{Q}}(i, j) = \|\mathbf{x}_i\| [\|\mathbf{x}_j\| \cos \theta_{\mathbf{x}_i \mathbf{x}_j}],$$

where the signed magnitude of the vector projection of  $\mathbf{x}_j$  along the axis of  $\mathbf{x}_i$

$$\begin{aligned} \text{comp}_{\vec{\mathbf{x}}_i}(\vec{\mathbf{x}}_j) &= \|\mathbf{x}_j\| \cos \theta_{\mathbf{x}_i \mathbf{x}_j} \\ &= \left( \frac{\mathbf{x}_i}{\|\mathbf{x}_i\|} \right)^T \mathbf{x}_j, \end{aligned}$$

provides a measure of the first-degree components (point coordinates) of the vector  $\mathbf{x}_j$

$$\mathbf{x}_j = (x_{j1}, x_{j2}, \dots, x_{jd})^T,$$

along the axis of the vector  $\mathbf{x}_i$

$$\mathbf{x}_i = (x_{i_1}, x_{i_2}, \dots, x_{i_d})^T.$$

Note that  $\text{comp}_{\vec{\mathbf{x}}_i}(\vec{\mathbf{x}}_j)$  is positive or negative if  $0 < \theta \leq \frac{\pi}{2}$  or  $\frac{\pi}{2} < \theta \leq \pi$  respectively. Also, if  $\theta = \frac{\pi}{2}$ , then  $\text{comp}_{\vec{\mathbf{x}}_i}(\vec{\mathbf{x}}_j) = 0$ .

Using the above assumptions and notation, given any row  $\tilde{\mathbf{Q}}(i, :)$  of Eq. (47), it follows that the statistic denoted by  $E_{\mathbf{x}_i}[\mathbf{x}_i | \{\mathbf{x}_j\}_{j=1}^N]$

$$\begin{aligned} E_{\mathbf{x}_i}[\mathbf{x}_i | \{\mathbf{x}_j\}_{j=1}^N] &= \|\mathbf{x}_i\| \sum_j \text{comp}_{\vec{\mathbf{x}}_i}(\vec{\mathbf{x}}_j) \\ &= \|\mathbf{x}_i\| \sum_j \|\mathbf{x}_j\| \cos \theta_{\mathbf{x}_i \mathbf{x}_j}, \end{aligned} \quad (48)$$

provides an estimate  $E_{\mathbf{x}_i}[\mathbf{x}_i | \{\mathbf{x}_j\}_{j=1}^N]$  for the amount of first degree components of the vector  $\mathbf{x}_i$  that are contained in a set of training vectors  $\{\mathbf{x}_j\}_{j=1}^N$ , where training labels have not been taken into account. It is concluded that Eq. (48) describes a distribution of first degree coordinates for the pattern vector  $\mathbf{x}_i$  in a training data collection.

Given that Eq. (48) involves signed magnitudes of vector projections along the axis of a fixed vector  $\mathbf{x}_i$ , the distribution of first degree vector coordinates described by Eq. (48) is said to *determine a first order statistical moment about the geometric locus of a data point  $\mathbf{x}_i$* . Because the statistic  $E_{\mathbf{x}_i}[\mathbf{x}_i | \{\mathbf{x}_j\}_{j=1}^N]$  depends on the uniform direction of  $\mathbf{x}_i$ , the statistic  $E_{\mathbf{x}_i}[\mathbf{x}_i | \{\mathbf{x}_j\}_{j=1}^N]$  is said to be unidirectional.

Alternatively, element  $\tilde{\mathbf{Q}}(i, j)$  in the  $j$ th column  $\tilde{\mathbf{Q}}(:, j)$  of Eq. (47) encodes the length of the vector  $\mathbf{x}_i$  multiplied by the scalar projection of  $\mathbf{x}_j$  onto  $\mathbf{x}_i$

$$\tilde{\mathbf{Q}}(i, j) = \|\mathbf{x}_i\| [\|\mathbf{x}_j\| \cos \theta_{\mathbf{x}_i \mathbf{x}_j}],$$

where the signed magnitude of the vector projection of  $\mathbf{x}_j$ , along each axis of a given training vector  $\mathbf{x}_i$ , provides an estimate of how much of the first degree components of the training vector  $\mathbf{x}_i$  are contained in the vector  $\mathbf{x}_j$ . It follows that the statistic denoted by  $E_{\mathbf{x}_j}[\mathbf{x}_j | \{\mathbf{x}_i\}_{i=1}^N]$

$$\begin{aligned} E_{\mathbf{x}_j}[\mathbf{x}_j | \{\mathbf{x}_i\}_{i=1}^N] &= \|\mathbf{x}_j\| \sum_i \text{comp}_{\vec{\mathbf{x}}_j}(\vec{\mathbf{x}}_i), \\ &= \sum_i \|\mathbf{x}_i\| \|\mathbf{x}_j\| \cos \theta_{\mathbf{x}_i \mathbf{x}_j}, \\ &= \|\mathbf{x}_j\| \sum_i \|\mathbf{x}_i\| \cos \theta_{\mathbf{x}_i \mathbf{x}_j}, \end{aligned} \quad (49)$$

provides an estimate  $E_{\mathbf{x}_j}[\mathbf{x}_j | \{\mathbf{x}_i\}_{i=1}^N]$  for the amount of first degree vector coordinates of a training data collection  $\{\mathbf{x}_i\}_{i=1}^N$  that are contained in the pattern vector  $\mathbf{x}_j$ , where training labels have not been taken into account. Because

the statistic  $E_{\mathbf{x}_j} [\mathbf{x}_j | \{\mathbf{x}_i\}_{i=1}^N]$  depends on the directions of all of the training vectors of  $\{\mathbf{x}_i\}_{i=1}^N$ , the statistic  $E_{\mathbf{x}_j} [\mathbf{x}_j | \{\mathbf{x}_i\}_{i=1}^N]$  is said to be omnidirectional.

It is concluded that row  $\tilde{\mathbf{Q}}(i, :)$  of Eq. (47) encodes distributions of first degree vector coordinates for a training vector  $\mathbf{x}_i$  within a training data collection  $\{\mathbf{x}_j\}_{j=1}^N$ , and that column  $\tilde{\mathbf{Q}}(:, j)$  of Eq. (47) encodes distributions of first degree vector coordinates for a training data collection  $\{\mathbf{x}_i\}_{i=1}^N$  within a training vector  $\mathbf{x}_j$ . The next section will develop unidirectional (pointwise) covariance statistics, which encode distributions of first degree vector coordinates for individual pattern vectors  $\mathbf{x}_i$  within training data collections  $\{\mathbf{x}_j\}_{j=1}^N$ .

## 12.3 Omnidirectional and Unidirectional Covariance Statistics

It will first be argued that classical covariance statistics provide omnidirectional, and therefore non-coherent, estimates of the joint variation of the random variables of a collection of training vectors about their common mean. Pointwise covariance statistics will then be developed. Pointwise covariance statistics provide a unidirectional estimate of how much a group of data and their common mean varies from a given vector, where the axis of the given vector is a fixed reference axis. Omnidirectional covariance statistics are considered next.

### 12.3.1 Omnidirectional Covariance Statistics

Take the data matrix  $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N)^T$  and consider the classical covariance statistic:

$$\begin{aligned} \widehat{\text{cov}}(\mathbf{X}) &= \frac{1}{N} \sum_i (\mathbf{x}_i - \bar{\mathbf{x}})^2, \\ &= \frac{1}{N} \sum_i \left( \mathbf{x}_i - \left( \frac{1}{N} \sum_i \mathbf{x}_i \right) \right)^2, \end{aligned} \quad (50)$$

written in vector notation. The statistic  $\widehat{\text{cov}}(\mathbf{X})$  measures the Euclidean distance between a common mean vector  $\bar{\mathbf{x}}$  and each of the training vectors  $\mathbf{x}_i$  in a collection of training data  $\{\mathbf{x}_i\}_{i=1}^N$  Ash [1993], Flury [1997]. Because the statistic  $\widehat{\text{cov}}(\mathbf{X})$  depends on  $N$  directions of  $N$  training vectors, the statistic  $\widehat{\text{cov}}(\mathbf{X})$  is said to be omnidirectional. The statistic  $\widehat{\text{cov}}(\mathbf{X})$  provides an omnidirectional estimate of the joint variation of the  $d \times N$  random variables of a collection of  $N$  pattern vectors  $\{\mathbf{x}_i\}_{i=1}^N$  about the geometric locus of the mean vector  $\bar{\mathbf{x}}$ , where training labels are not taken into account.

The statistic  $\widehat{\text{cov}}(\mathbf{X})$  in Eq. (50) produces a scalar quantity of a covariance estimate. A statistic is now developed that produces a vector quantity of a covariance estimate, where the statistic encodes a magnitude and a direction. The statistic provides a measure of how much a group of data and its common mean varies from a given vector, where the measure involves signed magnitudes of vector projections along the axis of the given vector. More specifically, a

pointwise covariance statistic  $\widehat{\text{cov}}_{up}(\mathbf{x}_i)$  provides a unidirectional estimate of how much a group of data  $\{\mathbf{x}_j\}_{j=1}^N$  and its common mean  $\bar{\mathbf{x}}$  varies from a given vector  $\mathbf{x}_i$ , where the axis of the vector  $\mathbf{x}_i$  is a fixed reference axis, and the Euclidean distance  $\|\mathbf{x}_i\| \|\mathbf{x}_j\| \cos \theta_{\mathbf{x}_i \mathbf{x}_j}$  between  $\mathbf{x}_i$  and each of the training vectors  $\{\mathbf{x}_j\}_{j=1}^N$  encodes the signed magnitude of the vector projection

$$\|\mathbf{x}_j\| \cos \theta_{\mathbf{x}_i \mathbf{x}_j},$$

along the axis of  $\mathbf{x}_i$ , where  $\theta_{\mathbf{x}_i \mathbf{x}_j}$  is the angle between  $\mathbf{x}_i$  and  $\mathbf{x}_j$ . Likewise, the Euclidean distance  $\|\mathbf{x}_i\| \|\bar{\mathbf{x}}\| \cos \theta_{\mathbf{x}_i \bar{\mathbf{x}}}$  between  $\mathbf{x}_i$  and the mean vector  $\bar{\mathbf{x}}$  encodes the signed magnitude of the vector projection

$$\|\bar{\mathbf{x}}\| \cos \theta_{\mathbf{x}_i \bar{\mathbf{x}}},$$

along the axis of  $\mathbf{x}_i$ , where  $\cos \theta_{\mathbf{x}_i \bar{\mathbf{x}}}$  is the angle between  $\mathbf{x}_i$  and  $\bar{\mathbf{x}}$ . Because the statistic  $\text{cov}_{up}(\mathbf{x}_i)$  depends on the uniform direction of  $\mathbf{x}_i$ , the statistic  $\text{cov}_{up}(\mathbf{x}_i)$  is said to be unidirectional.

Pointwise covariance statistics  $\widehat{\text{cov}}_{up}(\mathbf{x}_i)$  are now developed which are shown to determine first and second order statistical moments about the geometric loci of individual training vectors.

### 12.3.2 Pointwise Covariance Statistics

Take any row  $\tilde{\mathbf{Q}}(i, :)$  of the matrix  $\tilde{\mathbf{Q}}$  in Eq. (47) and consider the inner product statistic  $\|\mathbf{x}_i\| \|\mathbf{x}_j\| \cos \theta_{\mathbf{x}_i \mathbf{x}_j}$  in element  $\tilde{\mathbf{Q}}(i, j)$ . Given Eqs (8) and (11), it follows that element  $\tilde{\mathbf{Q}}(i, j)$  in row  $\tilde{\mathbf{Q}}(i, :)$  encodes the joint variation  $\text{cov}(\mathbf{x}_i, \mathbf{x}_j)$

$$\text{cov}(\mathbf{x}_i, \mathbf{x}_j) = \|\mathbf{x}_i\| \|\mathbf{x}_j\| \cos \theta_{\mathbf{x}_i \mathbf{x}_j},$$

between the vector components (point coordinates) of the geometric locus of the vector  $\mathbf{x}_i$

$$\left( \|\mathbf{x}_i\| \cos \alpha_{\mathbf{x}_i 1}, \quad \|\mathbf{x}_i\| \cos \alpha_{\mathbf{x}_i 2}, \quad \dots, \quad \|\mathbf{x}_i\| \cos \alpha_{\mathbf{x}_i d} \right),$$

and the vector components (point coordinates) of the geometric locus of the vector  $\mathbf{x}_j$

$$\left( \|\mathbf{x}_j\| \cos \alpha_{\mathbf{x}_j 1}, \quad \|\mathbf{x}_j\| \cos \alpha_{\mathbf{x}_j 2}, \quad \dots, \quad \|\mathbf{x}_j\| \cos \alpha_{\mathbf{x}_j d} \right),$$

so that the  $j$ th element  $\tilde{\mathbf{Q}}(i, j)$  of row  $\tilde{\mathbf{Q}}(i, :)$  encodes the joint variation of the  $d$  variables of a training vector  $\mathbf{x}_j$  about the  $d$  variables of the training vector  $\mathbf{x}_i$ . Thus, row  $\tilde{\mathbf{Q}}(i, :)$  encodes the joint variations between a vector  $\mathbf{x}_i$  and an entire collection of training data.

Again, take any row  $\tilde{\mathbf{Q}}(i, :)$  of the matrix  $\tilde{\mathbf{Q}}$  in Eq. (47). Given Eq. (12), it

follows that the statistic  $\widehat{\text{cov}}_{up}(\mathbf{x}_i)$ :

$$\begin{aligned}\widehat{\text{cov}}_{up}(\mathbf{x}_i) &= \sum_{j=1}^N \|\mathbf{x}_i\| \|\mathbf{x}_j\| \cos \theta_{\mathbf{x}_i \mathbf{x}_j}, \\ &= \sum_{j=1}^N \mathbf{x}_i^T \mathbf{x}_j, \\ &= \mathbf{x}_i^T \left( \sum_{j=1}^N \mathbf{x}_j \right), \\ &= \|\mathbf{x}_i\| \sum_{j=1}^N \|\mathbf{x}_j\| \cos \theta_{\mathbf{x}_i \mathbf{x}_j},\end{aligned}\tag{51}$$

provides a unidirectional estimate of the joint variation of the  $d$  variables of each of the  $N$  training vectors of a training data collection  $\{\mathbf{x}_j\}_{j=1}^N$  and the  $d$  variables of the common mean  $\sum_{j=1}^N \mathbf{x}_j$  of the training data, about the  $d$  variables of the vector  $\mathbf{x}_i$ , along the axis of  $\mathbf{x}_i$ . Note that Eq. (51) does not take training labels into account. The statistic  $\widehat{\text{cov}}_{up}(\mathbf{x}_i)$  encodes the direction of the vector  $\mathbf{x}_i$  and a signed magnitude along the axis of  $\mathbf{x}_i$ .

The statistic  $\widehat{\text{cov}}_{up}(\mathbf{x}_i)$  in Eq. (51) is defined to be a pointwise covariance estimate for the data point  $\mathbf{x}_i$ , where the statistic  $\widehat{\text{cov}}_{up}(\mathbf{x}_i)$  provides a unidirectional estimate of the joint variations between the geometric locus of each training vector  $\mathbf{x}_j$  and the geometric locus of the vector  $\mathbf{x}_i$ , which includes a unidirectional estimate of the joint variations between the locus of the mean vector  $\sum_{j=1}^N \mathbf{x}_j$  and the locus of the given vector  $\mathbf{x}_i$ . Given that the joint variations estimated by the statistic  $\widehat{\text{cov}}_{up}(\mathbf{x}_i)$  are derived from second order distance statistics  $\|\mathbf{x}_i - \mathbf{x}_j\|^2$ , which involve signed magnitudes of vector projections along the common axis of the vector  $\mathbf{x}_i$ , a pointwise covariance estimate  $\widehat{\text{cov}}_{up}(\mathbf{x}_i)$  is said to determine a *second order statistical moment about the geometric locus of the data point  $\mathbf{x}_i$* .

Returning to Eq. (48), it follows that Eq. (51) also encodes a distribution of first order coordinates for the training vector  $\mathbf{x}_i$ , which determines a first order statistical moment about the geometric locus of  $\mathbf{x}_i$ . The distribution of first order coordinates for  $\mathbf{x}_i$  describes how the components of  $\mathbf{x}_i$  are distributed within a training data collection. It is concluded that Eq. (51) determines a first and second order statistical moment about the geometric locus of the training point  $\mathbf{x}_i$ . It will now be demonstrated how pointwise covariance statistics can be used to find extreme data points which possess large pointwise covariances.

## 12.4 Discovery of Extreme Data Points with Pointwise Covariance Statistics

The Gramian matrix associated with the constrained quadratic form in Eq. (40) encodes inner product statistics for two labeled collections of training data. Denote those data points that belong to pattern class  $\mathbf{X}_1$  by  $\mathbf{x}_{1i}$  and those that belong to pattern class  $\mathbf{X}_2$  by  $\mathbf{x}_{2i}$ . Let  $\bar{\mathbf{x}}_1$  and  $\bar{\mathbf{x}}_2$  denote the mean vectors of pattern class  $\mathbf{X}_1$  and pattern class  $\mathbf{X}_2$  respectively. Let  $i = 1 : n_1$  where the

pattern vector  $\mathbf{x}_{1_i}$  has the training label  $y_i = 1$ , and let  $i = n_1 + 1 : n_2$  where the pattern vector  $\mathbf{x}_{2_i}$  has the training label  $y_i = -1$ . Using training label information, Eq. (51) can be rewritten as

$$\widehat{\text{cov}}_{up}(\mathbf{x}_{1_i}) = \mathbf{x}_{1_i}^T \left( \sum_{j=1}^{n_1} \mathbf{x}_{1_j} - \sum_{j=n_1+1}^{n_2} \mathbf{x}_{2_j} \right),$$

and

$$\widehat{\text{cov}}_{up}(\mathbf{x}_{2_i}) = \mathbf{x}_{2_i}^T \left( \sum_{j=n_1+1}^{n_2} \mathbf{x}_{2_j} - \sum_{j=1}^{n_1} \mathbf{x}_{1_j} \right).$$

It will now be shown that extreme training points possess large pointwise covariances relative to the non-extreme training points in each respective pattern class. Denote an extreme training point by  $\mathbf{x}_{1_{i*}}$  or  $\mathbf{x}_{2_{i*}}$  and a non-extreme training point by  $\mathbf{x}_{1_i}$  or  $\mathbf{x}_{2_i}$ .

Take any extreme training vector  $\mathbf{x}_{1_{i*}}$  and any non-extreme training vector  $\mathbf{x}_{1_i}$  that belong to the  $\mathbf{X}_1$  pattern class and consider the pointwise covariance estimates of the extreme data point  $\mathbf{x}_{1_{i*}}$ :

$$\begin{aligned} \widehat{\text{cov}}_{up}(\mathbf{x}_{1_{i*}}) &= \mathbf{x}_{1_{i*}}^T \left( \sum_{j=1}^{n_1} \mathbf{x}_{1_j} - \sum_{j=n_1+1}^{n_2} \mathbf{x}_{2_j} \right), \\ &= \mathbf{x}_{1_{i*}}^T \bar{\mathbf{x}}_1 - \mathbf{x}_{1_{i*}}^T \bar{\mathbf{x}}_2, \end{aligned}$$

and the non-extreme data point  $\mathbf{x}_{1_i}$ :

$$\begin{aligned} \widehat{\text{cov}}_{up}(\mathbf{x}_{1_i}) &= \mathbf{x}_{1_i}^T \left( \sum_{j=1}^{n_1} \mathbf{x}_{1_j} - \sum_{j=n_1+1}^{n_2} \mathbf{x}_{2_j} \right), \\ &= \mathbf{x}_{1_i}^T \bar{\mathbf{x}}_1 - \mathbf{x}_{1_i}^T \bar{\mathbf{x}}_2. \end{aligned}$$

Given that  $\mathbf{x}_{1_{i*}}$  is an extreme data point, it follows that  $\mathbf{x}_{1_{i*}}^T \bar{\mathbf{x}}_1 > \mathbf{x}_{1_i}^T \bar{\mathbf{x}}_1$  and that  $\mathbf{x}_{1_{i*}}^T \bar{\mathbf{x}}_2 < \mathbf{x}_{1_i}^T \bar{\mathbf{x}}_2$ , which shows that  $\widehat{\text{cov}}_{up}(\mathbf{x}_{1_{i*}}) > \widehat{\text{cov}}_{up}(\mathbf{x}_{1_i})$ .

Now take any extreme training vector  $\mathbf{x}_{2_{i*}}$  and any non-extreme training vector  $\mathbf{x}_{2_i}$  that belong to the  $\mathbf{X}_2$  pattern class and consider the pointwise covariance estimates of the extreme data point  $\mathbf{x}_{2_{i*}}$ :

$$\begin{aligned} \widehat{\text{cov}}_{up}(\mathbf{x}_{2_{i*}}) &= \mathbf{x}_{2_{i*}}^T \left( \sum_{j=n_1+1}^{n_2} \mathbf{x}_{2_j} - \sum_{j=1}^{n_1} \mathbf{x}_{1_j} \right), \\ &= \mathbf{x}_{2_{i*}}^T \bar{\mathbf{x}}_2 - \mathbf{x}_{2_{i*}}^T \bar{\mathbf{x}}_1, \end{aligned}$$

and the non-extreme data point  $\mathbf{x}_{2_i}$ :

$$\begin{aligned} \widehat{\text{cov}}_{up}(\mathbf{x}_{2_i}) &= \mathbf{x}_{2_i}^T \left( \sum_{j=n_1+1}^{n_2} \mathbf{x}_{2_j} - \sum_{j=1}^{n_1} \mathbf{x}_{1_j} \right), \\ &= \mathbf{x}_{2_i}^T \bar{\mathbf{x}}_2 - \mathbf{x}_{2_i}^T \bar{\mathbf{x}}_1. \end{aligned}$$

Given that  $\mathbf{x}_{2_{i*}}$  is an extreme data point, it follows that  $\mathbf{x}_{2_{i*}}^T \bar{\mathbf{x}}_2 > \mathbf{x}_{2_i}^T \bar{\mathbf{x}}_2$  and that  $\mathbf{x}_{2_{i*}}^T \bar{\mathbf{x}}_1 < \mathbf{x}_{2_i}^T \bar{\mathbf{x}}_1$ , which shows that  $\widehat{\text{cov}}_{up}(\mathbf{x}_{2_{i*}}) > \widehat{\text{cov}}_{up}(\mathbf{x}_{2_i})$ .

It is concluded that extreme training points possess large pointwise covariances relative to the non-extreme training points in their respective pattern class.

## 12.5 Eigen-scaled Pointwise Covariance Statistics

Consider again the Gramian matrix associated with the constrained quadratic form in Eq. (40) which encodes inner product statistics for two labeled collections of training data. This section will define eigen-scaled, pointwise covariance statistics for two collections of labeled training data. Denote those data points that belong to pattern class  $\mathbf{X}_1$  by  $\mathbf{x}_{1_i}$  and those that belong to pattern class  $\mathbf{X}_2$  by  $\mathbf{x}_{2_i}$ . Let  $\bar{\mathbf{x}}_1$  and  $\bar{\mathbf{x}}_2$  denote the mean vectors of pattern class  $\mathbf{X}_1$  and pattern class  $\mathbf{X}_2$  respectively. Let  $i = 1 : n_1$  where the pattern vector  $\mathbf{x}_{1_i}$  has the training label  $y_i = 1$  and let  $i = n_1 + 1 : n_2$  where the pattern vector  $\mathbf{x}_{2_i}$  has the training label  $y_i = -1$ .

Suppose that a principal eigen-decomposition of  $\mathbf{Q}$  provides two distinct types of eigen-scales  $\psi_{1j}$  and  $\psi_{2j}$  for the column vectors  $\mathbf{Q}(:, j)$  of  $\mathbf{Q}$  which define eigen-scales for elements in the rows  $\mathbf{Q}(i, :)$  of  $\mathbf{Q}$ , where the eigen-scales denoted by  $\psi_{1j}$  are correlated with the pattern class  $\mathbf{X}_1$ , the eigen-scales denoted by  $\psi_{2j}$  are correlated with the pattern class  $\mathbf{X}_2$ , and not all of the eigen-scales exceed zero.

Given Eq. (51) and the notation and assumptions outlined above, it follows that summation over the eigen-scaled  $\psi_{1j}$  and  $\psi_{2j}$  elements of row  $\mathbf{Q}(i, :)$  denoted in Eq. (47) provides eigen-scaled pointwise covariance estimates for the training vectors  $\mathbf{x}_{1_i}$ :

$$\begin{aligned} \widehat{\text{cov}}_{up}(\mathbf{x}_{1_i}) &= \|\mathbf{x}_{1_i}\| \sum_{j=1}^{n_1} \psi_{1j} \|\mathbf{x}_{1_j}\| \cos \theta_{\mathbf{x}_{1_i} \mathbf{x}_{1_j}} \\ &\quad - \|\mathbf{x}_{1_i}\| \sum_{j=n_1+1}^{n_2} \psi_{2j} \|\mathbf{x}_{2_j}\| \cos \theta_{\mathbf{x}_{1_i} \mathbf{x}_{2_j}}, \end{aligned} \quad (52)$$

and  $\mathbf{x}_{2_i}$ :

$$\begin{aligned} \widehat{\text{cov}}_{up}(\mathbf{x}_{2_i}) &= \|\mathbf{x}_{2_i}\| \sum_{j=n_1+1}^{n_2} \psi_{2j} \|\mathbf{x}_{2_j}\| \cos \theta_{\mathbf{x}_{2_i} \mathbf{x}_{2_j}} \\ &\quad - \|\mathbf{x}_{2_i}\| \sum_{j=1}^{n_1} \psi_{1j} \|\mathbf{x}_{1_j}\| \cos \theta_{\mathbf{x}_{2_i} \mathbf{x}_{1_j}}. \end{aligned} \quad (53)$$

Given Eqs (38) and (51), it follows that Eqs (52) and (53) determine eigen-balanced first and second order statistical moments about the geometric locus of a training point.

It has been demonstrated that extreme training points possess large pointwise covariances relative to the non-extreme training points in their respective pattern class. Therefore, it will be assumed that any given extreme data point  $\mathbf{x}_{1_{i_*}}$  or  $\mathbf{x}_{2_{i_*}}$  exhibits a critical first and second order statistical moment that exceeds some threshold for which  $\psi_{1_{i_*}} > 0$  or  $\psi_{2_{i_*}} > 0$ . This implies that first and second order statistical moments  $\widehat{\text{cov}}_{up}(\mathbf{x}_{1_i})$  and  $\widehat{\text{cov}}_{up}(\mathbf{x}_{2_i})$  about the loci of non-extreme data points  $\mathbf{x}_{1_i}$  and  $\mathbf{x}_{2_i}$  do not exceed the threshold, so that  $\psi_{1_i} = 0$  and  $\psi_{2_i} = 0$ . This indicates that the eigen-scaled pointwise covariance estimates in Eqs (52) and (53) are a function of eigen-scaled extreme training points. The next section will consider descriptive statistics for separating hyperplanes.



## 12.6 Descriptive Statistics for Separating Hyperplanes

It has previously been argued that the inner product statistics encoded within a matrix associated with a quadratic form determine the geometric shapes of second-order surfaces. It will now be argued that inner product statistics which have the algebraic form of  $\mathbf{x}_i^T \mathbf{x}_j$  describe hyperplane surfaces.

### The Coordinate Equation Version of a Hyperplane Surface

Every equation of the first degree specifies the locus of a linear curve or surface Nichols [1893], Tanner and Allen [1898], Eisenhart [1939]. Indeed, the coordinate equation version of an  $N$ -dimensional hyperplane surface:

$$Ax_{i_1} + Bx_{i_2} + \dots + Nx_{i_N} = P,$$

contains  $N$  first degree vector coordinates  $x_{i_j}$ . Now take the Gram matrix  $\mathbf{Q}$  associated with the constrained quadratic form in Eq. (40). Given that (1) the elements  $\mathbf{x}_i^T \mathbf{x}_j$  of  $\mathbf{Q}$  describe the geometric shapes of three, symmetric quadratic partitioning surfaces, and that (2) any row  $\mathbf{Q}(i, :)$  of  $\mathbf{Q}$  encodes a distribution of first degree vector coordinates for a training vector  $\mathbf{x}_i$ , it follows that  $\mathbf{Q}$  contains descriptive statistics for three, symmetrical hyperplane partitioning surfaces. It has been demonstrated by simulation studies that inner product statistics, which have the algebraic form of  $\mathbf{x}_i^T \mathbf{x}_j$ , do indeed provide descriptive statistics for three, symmetrical hyperplane partitioning surfaces Reeves [2007], Reeves [2009], Reeves and Jacyna [2011].

For the analyses that follow, it will be assumed that the Gram matrix  $\mathbf{Q}$  associated with the constrained quadratic form in Eq. (40) describes three, symmetrical hyperplane partitioning surfaces. Equations (52) and (53) will be used, in connection with a principal eigen-decomposition of the Gram matrix  $\mathbf{Q}$  associated with the constrained quadratic form in Eq. (40), to demonstrate how each Wolfe dual normal eigenaxis component on  $\psi$  encodes an eigen-balanced, point-wise covariance estimate for an extreme data point, which determines a properly proportioned eigen-scale that determines a suitable magnitude, and therefore a suitable eigenlocus (location), for a constrained primal normal eigenaxis component on  $\tau$ . Thereby, the analyses will demonstrate how the eigenlocus and corresponding eigenstate of each eigen-scaled extreme point on  $\tau$  is determined by the eigenlocus of a Wolfe dual normal eigenaxis component. Later on, an expression will be derived for the total allowed eigenenergy of an eigen-scaled extreme point and its corresponding eigenstate. The total allowed eigenenergies of a strong dual normal eigenlocus are revisited next.

## 12.7 Symmetrical Relationships Between the Total Allowed Eigenenergies of a Strong Dual Normal Eigenlocus

It has previously been claimed that strong duality relationships between the algebraic systems of  $\min \Psi(\tau)$  and  $\max \Xi(\psi)$  impose some type of symmetrical

relationships between the total allowed eigenenergies of  $\tau$  and  $\psi$

$$\|\tau\|_{\min_c}^2 \cong \|\psi\|_{\min_c}^2.$$

It has also been claimed that the lengths of the Wolfe dual normal eigenaxis components must be selected so that the total allowed eigenenergies of  $\tau_1$  and  $\tau_2$  are balanced by means of a symmetric equalizer statistic  $\nabla_{eq}$

$$\left(\|\tau_1\|_{\min_c}^2 - \|\tau_1\| \|\tau_2\| \cos \theta_{\tau_1 \tau_2}\right) + \nabla_{eq} \Leftrightarrow \left(\|\tau_2\|_{\min_c}^2 - \|\tau_1\| \|\tau_2\| \cos \theta_{\tau_1 \tau_2}\right) - \nabla_{eq},$$

in relation to a centrally located statistical fulcrum  $f_s$ , whereby the principal statistical state  $\lambda_{\max_\psi}$  and the characteristic eigenstates

$$\{\Psi_{\tau_1}(\mathbb{R}^d)\}_{i=1}^{l_1} \triangleq \left\{ \psi_{1i*} \mathbf{x}_{1i*} | \mathbf{x}_{1i*} \sim p_{\mathbf{x}_{1i} | \mathbf{X}_1}(\mathbf{x}_{1i} | \mathbf{X}_1) \right\}_{i=1}^{l_1},$$

and

$$\{\Psi_{\tau_2}(\mathbb{R}^d)\}_{i=1}^{l_2} \triangleq \left\{ \psi_{2i*} \mathbf{x}_{2i*} | \mathbf{x}_{2i*} \sim p_{\mathbf{x}_{2i} | \mathbf{X}_2}(\mathbf{x}_{2i} | \mathbf{X}_2) \right\}_{i=1}^{l_2},$$

determine a point  $\tau$  of statistical equilibrium. Algebraic and statistical equations that determine the symmetrical relationships between the critical minimum eigenenergies  $\|\tau\|_{\min_c}^2$  and  $\|\psi\|_{\min_c}^2$  of  $\tau$  and  $\psi$  will be developed in Sections 14 - 17. The findings presented in Sections 14 and 15 will show how the geometric configuration of a constrained primal normal eigenlocus  $\tau$  is symmetrically shaped by the eigen-balanced vector components of a Wolfe dual normal eigenlocus  $\psi$ , all of which satisfy critical length constraints. Section 17 will demonstrate that the eigenenergies of the eigen-scaled extreme points on  $\tau_1$  and  $\tau_2$  are distributed in a symmetrical manner which symmetrically balances the critical minimum eigenenergies of  $\tau_1$  and  $\tau_2$ .

The paper will now motivate and introduce the form of an algebraic expression that will be used to examine eigenlocus relationships between the eigen-scaled extreme data points on  $\tau$  and the Wolfe dual normal eigenaxis components on  $\psi$ . Recall that the Wolfe dual normal eigenlocus  $\psi$  of Eq. (40)

$$\max \Xi(\psi) = \mathbf{1}^T \psi - \frac{\psi^T \mathbf{Q} \psi}{2},$$

which satisfies the inner product statistic  $\psi^T \mathbf{y} = 0$  and magnitude constraints  $\psi_i \geq 0$ , provides an estimate of the principal eigenvector  $\psi$  of the Gram matrix  $\mathbf{Q}$  associated with the constrained quadratic form  $\psi^T \mathbf{Q} \psi$ , whereby  $\psi$  is the principal eigenaxis of three, symmetrical hyperplane partitioning surfaces described by the constrained quadratic form  $\psi^T \mathbf{Q} \psi$ , and  $\mathbf{y}$  is a directrix which is orthogonal to  $\psi$ :  $\psi \perp \mathbf{y}$ . The strong duality relationships between the algebraic systems of  $\min \Psi(\tau)$  and  $\max \Xi(\psi)$  indicate that the principal eigenvector  $\psi$  of  $\mathbf{Q}$  satisfies a critical minimum eigenenergy constraint  $\|\psi\|_{\min_c}^2$  that is symmetrically related to the critical minimum eigenenergy constraint  $\|\tau\|_{\min_c}^2$  satisfied by a strong dual normal eigenlocus  $\tau$ .

Suppose that an expression is obtained for a principal eigen-decomposition of a Gram matrix  $\mathbf{Q}$

$$\max \mathbf{Q}\psi = \lambda_{\max_\psi} \psi,$$

which is associated with an estimate of a Wolfe dual normal eigenlocus  $\psi$ . Multiplying both sides of the above expression by the vector transpose  $\psi^T$  of a Wolfe dual normal eigenlocus  $\psi$  provides an expression which relates the constrained quadratic form  $\psi^T \mathbf{Q}\psi$  in Eq. (40) to the total allowed (critical minimum) eigenenergy  $\|\psi\|_{\min}^2$  of a Wolfe dual normal eigenlocus  $\psi$ :

$$\begin{aligned} \max \psi^T \mathbf{Q}\psi &= \psi^T \lambda_{\max_\psi} \psi, \\ &= \lambda_{\max_\psi} \psi^T \psi, \\ &= \lambda_{\max_\psi} \|\psi\|_{\min_c}^2, \end{aligned} \tag{54}$$

where the critical minimum eigenenergy  $\lambda_{\max_\psi} \|\psi\|_{\min}^2$  of  $\psi$  is symmetrically correlated with the critical minimum eigenenergy  $\|\tau\|_{\min}^2$  of a constrained primal normal eigenlocus  $\tau$

$$\max \psi^T \mathbf{Q}\psi \cong \|\tau\|_{\min_c}^2.$$

Given the algebraic relationships outlined directly above, this paper will examine the geometric and statistical properties exhibited by a Wolfe dual normal eigenlocus  $\psi$  by analyzing a principal eigen-decomposition of the Gram matrix  $\mathbf{Q}$  denoted in Eqs (46) and (47). The next part of the paper will examine a general expression of a principal eigen-decomposition that offers tractable point and coordinate relationships between the eigen-scaled extreme data points on  $\tau_1$  and  $\tau_2$  and their correlated Wolfe dual normal eigenaxis components on  $\psi$ . Analysis of the expression will provide significant insights into geometric and statistical interconnections between the constrained primal and the Wolfe dual normal eigenaxis components. Sections 14 and 15 will use the general expression for the principal eigen-decomposition to develop algebraic expressions for the eigenloci (the geometric locations) of the  $\psi_{1i*} \vec{\mathbf{e}}_{1i*}$  and the  $\psi_{2i*} \vec{\mathbf{e}}_{2i*}$  Wolfe dual normal eigenaxis components. These expressions will be used to define uniform geometric and statistical properties which are jointly exhibited by the Wolfe dual normal eigenaxis components on  $\psi$  and the constrained primal normal eigenaxis components on  $\tau$ . The general expression for the principal eigen-decomposition is obtained next.

### 13 Underneath the Hood of a Wolfe Dual Normal Eigenlocus

Take the Gram matrix  $\mathbf{Q}$  associated with the quadratic form in Eq (40). Let  $\mathbf{q}_j$  denote the  $j$ th column of  $\mathbf{Q}$ , which is an  $N$ -vector. Let  $\lambda_{\max_\psi}$  and  $\psi$  denote the largest eigenvalue and largest eigenvector of  $\mathbf{Q}$  respectively. Using this notation Trefethen and Bau [1998], the principal eigen-decomposition of  $\mathbf{Q}$ :

$$\mathbf{Q}\psi = \lambda_{\max_\psi} \psi,$$

can be rewritten as

$$\lambda_{\max_\psi} \psi = \sum_{j=1}^N \psi_j \mathbf{q}_j ,$$

where the Wolfe dual normal eigenaxis  $\psi$  of  $\mathbf{Q}$  is expressed as a linear combination of the eigen-transformed vectors  $\psi_j \mathbf{q}_j$  :

$$\lambda_{\max_\psi} \begin{bmatrix} \psi_1 \end{bmatrix} = \psi_1 \begin{bmatrix} \mathbf{q}_1 \end{bmatrix} + \psi_2 \begin{bmatrix} \mathbf{q}_2 \end{bmatrix} + \cdots + \psi_N \begin{bmatrix} \mathbf{q}_N \end{bmatrix} , \quad (55)$$

where the  $i$ th element of the vector  $\mathbf{q}_j$  encodes an inner product statistic  $\mathbf{x}_i^T \mathbf{x}_j$  between the vectors  $\mathbf{x}_i$  and  $\mathbf{x}_j$ .

Equation (55) is now used to examine how eigen-transformed, inner product statistical relationships between the extreme training vectors and the Wolfe dual normal eigenaxis components specify the eigenloci of the Wolfe dual normal eigenaxis components on  $\psi$ .

Using Eqs (46) and (55), a Wolfe dual normal eigenlocus  $\psi$

$$\psi = (\psi_1, \psi_2, \dots, \psi_N)^T ,$$

can be written as:

$$\begin{aligned} \psi = & \frac{\psi_1}{\lambda_{\max_\psi}} \begin{pmatrix} \mathbf{x}_1^T \mathbf{x}_1 \\ \mathbf{x}_2^T \mathbf{x}_1 \\ \vdots \\ -\mathbf{x}_N^T \mathbf{x}_1 \end{pmatrix} \\ & + \frac{\psi_2}{\lambda_{\max_\psi}} \begin{pmatrix} \mathbf{x}_1^T \mathbf{x}_2 \\ \mathbf{x}_2^T \mathbf{x}_2 \\ \vdots \\ -\mathbf{x}_N^T \mathbf{x}_2 \end{pmatrix} + \cdots \\ & \cdots + \frac{\psi_N}{\lambda_{\max_\psi}} \begin{pmatrix} -\mathbf{x}_1^T \mathbf{x}_N \\ -\mathbf{x}_2^T \mathbf{x}_N \\ \vdots \\ \mathbf{x}_N^T \mathbf{x}_N \end{pmatrix} , \end{aligned} \quad (56)$$

which illustrates that the magnitude  $\psi_j$  of the  $j^{th}$  Wolfe dual normal eigenaxis component is correlated with joint variations of the training data about the training vector  $\mathbf{x}_j$ .

Alternatively, using Eqs (47) and (55), a Wolfe dual normal eigenlocus  $\psi$

can be written as:

$$\begin{aligned} \psi = & \frac{\psi_1}{\lambda_{\max\psi}} \begin{pmatrix} \|\mathbf{x}_1\| \|\mathbf{x}_1\| \cos \theta_{\mathbf{x}_1^T \mathbf{x}_1} \\ \|\mathbf{x}_2\| \|\mathbf{x}_1\| \cos \theta_{\mathbf{x}_2^T \mathbf{x}_1} \\ \vdots \\ -\|\mathbf{x}_N\| \|\mathbf{x}_1\| \cos \theta_{\mathbf{x}_N^T \mathbf{x}_1} \end{pmatrix} + \dots \\ & \dots + \frac{\psi_N}{\lambda_{\max\psi}} \begin{pmatrix} -\|\mathbf{x}_1\| \|\mathbf{x}_N\| \cos \theta_{\mathbf{x}_1^T \mathbf{x}_N} \\ -\|\mathbf{x}_2\| \|\mathbf{x}_N\| \cos \theta_{\mathbf{x}_2^T \mathbf{x}_N} \\ \vdots \\ \|\mathbf{x}_N\| \|\mathbf{x}_N\| \cos \theta_{\mathbf{x}_N^T \mathbf{x}_N} \end{pmatrix}, \end{aligned} \quad (57)$$

which illustrates that the magnitude  $\psi_j$  of the  $j^{th}$  Wolfe dual normal eigenaxis component on  $\psi$  is correlated with scalar projections  $\|\mathbf{x}_j\| \cos \theta_{\mathbf{x}_i \mathbf{x}_j}$  of the training vector  $\mathbf{x}_j$  onto the training data.

### 13.1 Non-Orthogonal Eigenaxes of $\psi$

Express a Wolfe dual normal eigenlocus  $\psi$  in terms of  $l$  non-orthogonal unit vectors  $\{\vec{\mathbf{e}}_{1*}, \dots, \vec{\mathbf{e}}_{l*}\}$

$$\begin{aligned} \psi &= \sum_{i=1}^l \psi_{i*} \vec{\mathbf{e}}_{i*}, \\ &= \sum_{i=1}^{l_1} \psi_{1i*} \vec{\mathbf{e}}_{1i*} + \sum_{i=1}^{l_2} \psi_{2i*} \vec{\mathbf{e}}_{2i*}, \end{aligned} \quad (58)$$

where the eigen-scaled, non-orthogonal unit vector denoted by  $\psi_{1i*} \vec{\mathbf{e}}_{1i*}$  or  $\psi_{2i*} \vec{\mathbf{e}}_{2i*}$  is correlated with the extreme training vector  $\mathbf{x}_{1i*}$  or  $\mathbf{x}_{2i*}$  respectively. Accordingly, a Wolfe dual normal eigenaxis component  $\psi_{1i*} \vec{\mathbf{e}}_{1i*}$  or  $\psi_{2i*} \vec{\mathbf{e}}_{2i*}$  is an eigen-scaled, non-orthogonal unit vector that contributes to the estimation of  $\psi$ .

Equations (57) and (58) will be used to identify the geometric and statistical properties which determine the magnitudes and the directions of the Wolfe dual  $\psi_{1i*} \vec{\mathbf{e}}_{1i*}$  and  $\psi_{2i*} \vec{\mathbf{e}}_{2i*}$  and the magnitudes of the constrained primal  $\psi_{1i*} \mathbf{x}_{1i*}$  and  $\psi_{2i*} \mathbf{x}_{2i*}$  normal eigenaxis components. It will be demonstrated that each Wolfe dual normal eigenaxis component stores an eigen-balanced first and second order statistical moment about the locus of an extreme data point, which determines an eigen-scale for a constrained primal normal eigenaxis component, such that each correlated Wolfe dual and constrained primal normal eigenaxis component exhibit symmetrical magnitudes and symmetrical directions.

The next two sections will examine the geometric and statistical properties of the eigen-scales used to form the constrained primal normal eigenaxis components. It will be shown that the eigenlocus of each Wolfe dual normal eigenaxis component on  $\psi$  encodes an eigen-balanced, positive magnitude along the axis of a correlated extreme vector, which is determined by an eigen-balanced, first and second order statistical moment about the locus of the extreme vector,

such that the eigenlocus of each Wolfe dual normal eigenaxis component on  $\psi$  provides a symmetrical eigen-scale which determines the critical length of a correlated extreme training vector on  $\tau$ . The direction of each non-orthogonal unit vector  $\vec{e}_{1i*}$  or  $\vec{e}_{2i*}$  will be shown to be identical to the direction of an extreme training vector  $\mathbf{x}_{1i*}$  or  $\mathbf{x}_{2i*}$ . It will also be shown that the eigenloci of the constrained normal eigenaxis components on  $\psi$  and  $\tau$  delineate an implicit estimate of a separating hyperplane that is bounded by bilaterally symmetrical hyperplane borders. Sections 16, 17, and 18 will use results from the analysis to examine the geometric and statistical characteristics of the statistical equilibrium state implied by the KKT condition of Eq. (17). The analysis begins with extreme point notation and assumptions.

## Extreme Point Notation and Assumptions

Pattern category information defines a distinct pair of pattern classes. Denote the pattern classes one and two by  $\mathbf{X}_1$  and  $\mathbf{X}_2$  respectively. Denote those extreme points that belong to pattern class  $\mathbf{X}_1 = \{\mathbf{x}_{1i*} | \mathbf{x}_{1i*} \in \mathbf{X}_1, y_i = +1\}$  by  $\mathbf{x}_{1i*}$  and those that belong to pattern class  $\mathbf{X}_2 = \{\mathbf{x}_{2i*} | \mathbf{x}_{2i*} \in \mathbf{X}_2, y_i = -1\}$  by  $\mathbf{x}_{2i*}$ . Let  $l_1$  denote the number of extreme data points that belong to the pattern class  $\mathbf{X}_1$  and let  $l_2$  denote the number that belong to the pattern class  $\mathbf{X}_2$ . Let the extreme training point  $\mathbf{x}_{1i*}$  associated with the  $\psi_{1i*} \vec{e}_{1i*}$  normal eigenaxis component have the training label  $y_i = 1$ , and let the extreme training point  $\mathbf{x}_{2i*}$  associated with the  $\psi_{2i*} \vec{e}_{2i*}$  normal eigenaxis component have the training label  $y_i = -1$ . Denote the number of  $\psi_{1i*} \vec{e}_{1i*}$  and  $\psi_{2i*} \vec{e}_{2i*}$  normal eigenaxis components by  $l_1$  and  $l_2$  respectively. Assume that  $l_1 + l_2 = l$ .

Given the above assumptions and Eqs (52) and (53), it follows that an extreme data point  $\mathbf{x}_{1i*}$  possesses a large pointwise covariance

$$\begin{aligned} \widehat{\text{cov}}_{up\downarrow}(\mathbf{x}_{1i*}) &= \|\mathbf{x}_{1i*}\| \sum_{j=1}^{l_1} \psi_{1j*} \|\mathbf{x}_{1j*}\| \cos \theta_{\mathbf{x}_{1i*} \mathbf{x}_{1j*}} \\ &\quad - \|\mathbf{x}_{1i*}\| \sum_{j=1}^{l_2} \psi_{2j*} \|\mathbf{x}_{2j*}\| \cos \theta_{\mathbf{x}_{1i*} \mathbf{x}_{2j*}}, \end{aligned} \quad (59)$$

relative to the  $\mathbf{X}_1$  training vectors, where  $\widehat{\text{cov}}_{up\downarrow}(\mathbf{x}_{1i}) = 0$  for all non-extreme training vectors that belong to pattern class  $\mathbf{X}_1$ . Likewise, an extreme data point  $\mathbf{x}_{2i*}$  possesses a large pointwise covariance

$$\begin{aligned} \widehat{\text{cov}}_{up\downarrow}(\mathbf{x}_{2i*}) &= \|\mathbf{x}_{2i*}\| \sum_{j=1}^{l_2} \psi_{2j*} \|\mathbf{x}_{2j*}\| \cos \theta_{\mathbf{x}_{2i*} \mathbf{x}_{2j*}} \\ &\quad - \|\mathbf{x}_{2i*}\| \sum_{j=1}^{l_1} \psi_{1j*} \|\mathbf{x}_{1j*}\| \cos \theta_{\mathbf{x}_{2i*} \mathbf{x}_{1j*}}, \end{aligned} \quad (60)$$

relative to the  $\mathbf{X}_2$  training vectors, where  $\widehat{\text{cov}}_{up\downarrow}(\mathbf{x}_{2i}) = 0$  for all non-extreme training vectors that belong to pattern class  $\mathbf{X}_2$ .

The next section will examine the eigenloci of the  $\psi_{1i*} \vec{e}_{1i*}$  Wolfe dual normal eigenaxis components.

## 14 Eigenloci of the $\psi_{1i*} \vec{e}_{1i*}$ Wolfe Dual Normal Eigenaxis Components

Let  $i = 1 : l_1$ , where the extreme training vector  $\mathbf{x}_{1i*}$  has the training label  $y_i = 1$ . Using Eqs (57) and (58), the eigenlocus of the  $i^{th}$  Wolfe dual normal eigenaxis component  $\psi_{1i*} \vec{e}_{1i*}$  on  $\psi$  is a function of the expression:

$$\begin{aligned} \psi_{1i*} = & \lambda_{\max_\psi}^{-1} \|\mathbf{x}_{1i*}\| \sum_{j=1}^{l_1} \psi_{1j*} \|\mathbf{x}_{1j*}\| \cos \theta_{\mathbf{x}_{1i*} \mathbf{x}_{1j*}} \\ & - \lambda_{\max_\psi}^{-1} \|\mathbf{x}_{1i*}\| \sum_{j=1}^{l_2} \psi_{2j*} \|\mathbf{x}_{2j*}\| \cos \theta_{\mathbf{x}_{1i*} \mathbf{x}_{2j*}}, \end{aligned} \quad (61)$$

where  $\psi_{1i*}$  provides an eigen-scaling for the non-orthogonal unit vector  $\vec{e}_{1i*}$ . Geometric and statistical explanations for the eigenlocus statistics

$$\psi_{1j*} \|\mathbf{x}_{1j*}\| \cos \theta_{\mathbf{x}_{1i*} \mathbf{x}_{1j*}} \text{ and } \psi_{2j*} \|\mathbf{x}_{2j*}\| \cos \theta_{\mathbf{x}_{1i*} \mathbf{x}_{2j*}}, \quad (62)$$

in Eq. (61) are considered next.

### Geometric and Statistical Interpretations of $\psi_{1i*} \vec{e}_{1i*}$ Eigenlocus Statistics

The first geometric interpretation of the eigenlocus statistics in Eq. (62) defines the terms:

$$\psi_{1j*} \text{ and } \psi_{2j*},$$

to be eigen-scales for the signed magnitudes of the vector projections

$$\|\mathbf{x}_{1j*}\| \cos \theta_{\mathbf{x}_{1i*} \mathbf{x}_{1j*}} \text{ and } \|\mathbf{x}_{2j*}\| \cos \theta_{\mathbf{x}_{1i*} \mathbf{x}_{2j*}},$$

of the eigen-scaled extreme vectors  $\psi_{1j*} \mathbf{x}_{1j*}$  and  $\psi_{2j*} \mathbf{x}_{2j*}$  along the axis of the extreme vector  $\mathbf{x}_{1i*}$ , where  $\cos \theta_{\mathbf{x}_{1i*} \mathbf{x}_{1j*}}$  and  $\cos \theta_{\mathbf{x}_{1i*} \mathbf{x}_{2j*}}$  are the angles between the axes of the eigen-scaled extreme vectors  $\psi_{1j*} \mathbf{x}_{1j*}$  and  $\psi_{2j*} \mathbf{x}_{2j*}$  and the axis of the extreme vector  $\mathbf{x}_{1i*}$ . Figure 28 illustrates the geometric and statistical nature of the eigenlocus statistics in Eqs (62).

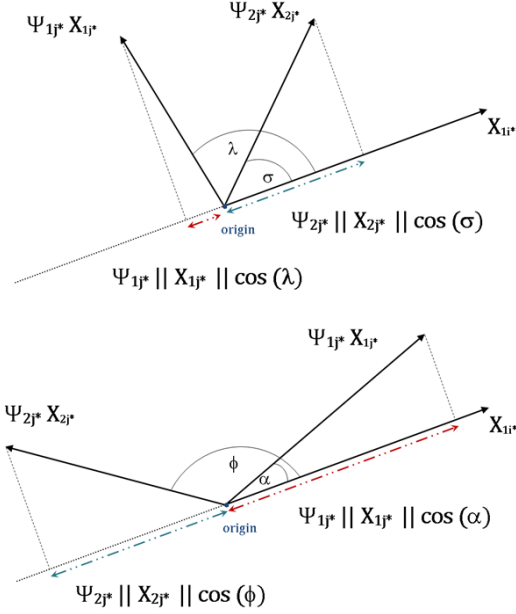


Figure 28: Illustration of eigen-scaled signed magnitudes of vector projections of eigen-scaled extreme vectors  $\psi_{1j*} \mathbf{x}_{1j*}$  and  $\psi_{2j*} \mathbf{x}_{2j*}$  along the axis of an extreme vector  $\mathbf{x}_{1i*}$  which is correlated with a Wolfe dual normal eigenaxis component  $\psi_{1i*} \vec{\mathbf{e}}_{1i*}$ . Any given eigen-scaled signed magnitude  $\psi_{1j*} \|\mathbf{x}_{1j*}\| \cos \theta_{\mathbf{x}_{1i*} \mathbf{x}_{1j*}}$  or  $\psi_{2j*} \|\mathbf{x}_{2j*}\| \cos \theta_{\mathbf{x}_{1i*} \mathbf{x}_{2j*}}$  may be positive or negative.

### An Alternative Geometric Interpretation

An alternative geometric explanation for the eigenlocus statistics in Eq. (62) accounts for the representation of the  $\tau_1$  and  $\tau_2$  primal normal eigenlocus components within the Wolfe dual eigenspace. Consider the algebraic relationships

$$\psi_{1j*} \|\mathbf{x}_{1j*}\| = \|\psi_{1j*} \mathbf{x}_{1j*}\| = \|\tau_1(j)\|,$$

and

$$\psi_{2j*} \|\mathbf{x}_{2j*}\| = \|\psi_{2j*} \mathbf{x}_{2j*}\| = \|\tau_2(j)\|,$$

where  $\tau_1(j)$  and  $\tau_2(j)$  are the  $j$ th constrained primal normal eigenaxis components on  $\tau_1$  and  $\tau_2$ . Given the above expressions, it follows that the eigen-scaled  $\psi_{1j*}$  signed magnitude  $\|\mathbf{x}_{1j*}\| \cos \theta_{\mathbf{x}_{1i*} \mathbf{x}_{1j*}}$  of the vector projection of the eigen-scaled extreme vector  $\psi_{1j*} \mathbf{x}_{1j*}$  along the axis of the extreme vector  $\mathbf{x}_{1i*}$

$$\psi_{1j*} \|\mathbf{x}_{1j*}\| \cos \theta_{\mathbf{x}_{1i*} \mathbf{x}_{1j*}},$$

encodes the cosine-scaled  $\cos \theta_{\mathbf{x}_{1i*} \mathbf{x}_{1j*}}$  length of the  $j$ th constrained primal normal eigenaxis component  $\tau_1(j)$  on  $\tau_1$

$$\cos \theta_{\mathbf{x}_{1i*} \mathbf{x}_{1j*}} \|\tau_1(j)\|,$$



where  $\psi_{1j*}$  is the magnitude of the  $\psi_{1j*} \vec{\mathbf{e}}_{1j*}$  Wolfe dual normal eigenaxis component, and  $\cos \theta_{\mathbf{x}_{1i*} \mathbf{x}_{1j*}}$  encodes the angle between the extreme training vectors  $\mathbf{x}_{1i*}$  and  $\mathbf{x}_{1j*}$ .

Likewise, the eigen-scaled  $\psi_{2j*}$  signed magnitude  $\|\mathbf{x}_{2j*}\| \cos \theta_{\mathbf{x}_{1i*} \mathbf{x}_{2j*}}$  of the vector projection of the eigen-scaled extreme vector  $\psi_{2j*} \mathbf{x}_{2j*}$  along the axis of the extreme vector  $\mathbf{x}_{1i*}$

$$\psi_{2j*} \|\mathbf{x}_{2j*}\| \cos \theta_{\mathbf{x}_{1i*} \mathbf{x}_{2j*}},$$

encodes the cosine-scaled  $\cos \theta_{\mathbf{x}_{1i*} \mathbf{x}_{2j*}}$  length of the  $j$ th constrained primal normal eigenaxis component  $\tau_2(j)$  on  $\tau_2$

$$\cos \theta_{\mathbf{x}_{1i*} \mathbf{x}_{2j*}} \|\tau_2(j)\|,$$

where  $\psi_{2j*}$  is the magnitude of the  $\psi_{2j*} \vec{\mathbf{e}}_{2j*}$  Wolfe dual normal eigenaxis component, and  $\cos \theta_{\mathbf{x}_{1i*} \mathbf{x}_{2j*}}$  encodes the angle between the extreme training vectors  $\mathbf{x}_{1i*}$  and  $\mathbf{x}_{2j*}$ .

Given the above analysis, it follows that the eigenlocus of the Wolfe dual normal eigenaxis component  $\psi_{1i*} \vec{\mathbf{e}}_{1i*}$  is a function of the constrained primal normal eigenaxis components on  $\tau_1$  and  $\tau_2$ :

$$\begin{aligned} \psi_{1i*} &= \lambda_{\max \psi}^{-1} \|\mathbf{x}_{1i*}\| \sum_{j=1}^{l_1} \cos \theta_{\mathbf{x}_{1i*} \mathbf{x}_{1j*}} \|\tau_1(j)\| \\ &\quad - \lambda_{\max \psi}^{-1} \|\mathbf{x}_{1i*}\| \sum_{j=1}^{l_2} \cos \theta_{\mathbf{x}_{1i*} \mathbf{x}_{2j*}} \|\tau_2(j)\|. \end{aligned} \quad (63)$$

Previous analyses and simulation studies indicate that the constrained primal normal eigenaxis components on  $\tau_1 - \tau_2$  account for a bipartite symmetric partitioning of a region of large covariance that is well-positioned between a pair of data distributions. The next analysis examines how uniform geometric and statistical properties which are jointly exhibited by the Wolfe dual  $\psi_{1i*} \vec{\mathbf{e}}_{1i*}$  and the constrained primal  $\psi_{1i*} \mathbf{x}_{1i*}$  normal eigenaxis components account for this bipartite symmetric partitioning.

#### 14.1 Uniform Geometric and Statistical Properties Jointly Exhibited by Normal Eigenaxis Components on $\psi$ and $\tau_1$

Using results from the previous analysis, the KKT constraint  $\sum_{i=1}^{l_1} \psi_{1i*} = \sum_{i=1}^{l_2} \psi_{2i*}$  of Eq. (38) indicates that Eq. (61) determines an eigen-balanced, signed magnitude along the axis of an extreme vector  $\mathbf{x}_{1i*}$ .

Let  $\text{comp}_{\mathbf{x}_{1i*}} \left( \overrightarrow{\tilde{\psi}_{1i*} \|\tilde{\mathbf{x}}_*\|_{1i*}} \right)$  denote the eigen-balanced, signed magnitude

$$\begin{aligned} \text{comp}_{\mathbf{x}_{1i*}} \left( \overrightarrow{\tilde{\psi}_{1i*} \|\tilde{\mathbf{x}}_*\|_{1i*}} \right) &= \sum_{j=1}^{l_1} \psi_{1j*} \\ &\quad \times \left[ \|\mathbf{x}_{1j*}\| \cos \theta_{\mathbf{x}_{1i*} \mathbf{x}_{1j*}} \right] \\ &\quad - \sum_{j=1}^{l_2} \psi_{2j*} \\ &\quad \times \left[ \|\mathbf{x}_{2j*}\| \cos \theta_{\mathbf{x}_{1i*} \mathbf{x}_{2j*}} \right], \end{aligned} \quad (64)$$

along the axis of the extreme vector  $\mathbf{x}_{1i*}$  that is correlated with the Wolfe dual normal eigenaxis component  $\psi_{1i*} \vec{\mathbf{e}}_{1i*}$ .

Given Eqs (38) and (59), it follows that Eq. (64) determines an eigen-balanced distribution of eigen-scaled first degree coordinates of extreme training vectors along the axis of  $\mathbf{x}_{1i*}$ ; Eqs (38) and (59) also indicate that Eq. (64) determines an eigen-balanced first and second order statistical moment about the geometric locus of  $\mathbf{x}_{1i*}$ .

Furthermore, Eqs (38), (61), and (64) show that joint distributions of the components of  $\psi$  and  $\tau$  are symmetrically distributed over the axis of the extreme vector  $\mathbf{x}_{1i*}$ . This indicates that joint distributions of the components of  $\psi$  and  $\tau$  are symmetrically distributed over the axis of the Wolfe dual normal eigenaxis component  $\psi_{1i*} \vec{\mathbf{e}}_{1i*}$ .

Alternatively, given Eq. (63), the eigen-balanced, signed magnitude in Eq. (64) depends upon the difference between integrated cosine-scaled magnitudes of the constrained primal normal eigenaxis components on  $\tau_1$  and  $\tau_2$ :

$$\begin{aligned} \text{comp}_{\mathbf{x}_{1i*}} \left( \overrightarrow{\tilde{\psi}_{1i*} \|\tilde{\mathbf{x}}_*\|_{1i*}} \right) &= \sum_{j=1}^{l_1} \cos \theta_{\mathbf{x}_{1i*} \mathbf{x}_{1j*}} \|\tau_1(j)\| \\ &\quad - \sum_{j=1}^{l_2} \cos \theta_{\mathbf{x}_{1i*} \mathbf{x}_{2j*}} \|\tau_2(j)\|, \end{aligned} \quad (65)$$

which also shows that joint distributions of the components of  $\psi$  and  $\tau$  are symmetrically distributed over the axes of both  $\psi_{1i*} \vec{\mathbf{e}}_{1i*}$  and  $\mathbf{x}_{1i*}$ .

Using Eqs (61) and (64), it follows that the length  $\psi_{1i*}$  of the Wolfe dual normal eigenaxis component  $\psi_{1i*} \vec{\mathbf{e}}_{1i*}$  is determined by a weighted length of the correlated extreme training vector  $\mathbf{x}_{1i*}$

$$\psi_{1i*} = \left[ \lambda_{\max_\psi}^{-1} \times \text{comp}_{\mathbf{x}_{1i*}} \left( \overrightarrow{\tilde{\psi}_{1i*} \|\tilde{\mathbf{x}}_*\|_{1i*}} \right) \right] \|\mathbf{x}_{1i*}\|, \quad (66)$$

where the weighting factor encodes an eigenvalue  $\lambda_{\max_\psi}^{-1}$  scaling of an eigen-balanced, signed magnitude  $\text{comp}_{\mathbf{x}_{1i*}} \left( \overrightarrow{\tilde{\psi}_{1i*} \|\tilde{\mathbf{x}}_*\|_{1i*}} \right)$  along the axis of  $\mathbf{x}_{1i*}$ .

Given that  $\psi_{1i*} > 0$ ,  $\lambda_{\max_\psi}^{-1} > 0$ , and  $\|\mathbf{x}_{1i*}\| > 0$ , it follows that the eigen-balanced, signed magnitude along the axis of  $\mathbf{x}_{1i*}$  is positive

$$\text{comp}_{\mathbf{x}_{1i*}} \left( \overrightarrow{\tilde{\psi}_{1i*} \|\tilde{\mathbf{x}}_*\|_{1i*}} \right) > 0,$$

which indicates that the weighting factor in Eq. (66) determines an eigen-balanced length

$$\lambda_{\max_\psi}^{-1} \text{comp}_{\mathbf{x}_{1i*}} \left( \overrightarrow{\tilde{\psi}_{1i*} \|\tilde{\mathbf{x}}_*\|_{1i*}} \right) \|\mathbf{x}_{1i*}\|,$$

for the extreme vector  $\mathbf{x}_{1i*}$ . Therefore, Eq. (66) determines an eigen-balanced length for both  $\psi_{1i*} \vec{\mathbf{e}}_{1i*}$  and  $\mathbf{x}_{1i*}$ .

Returning to Eqs (59) and (61), it follows that the length  $\psi_{1i*}$  of the Wolfe dual normal eigenaxis component  $\psi_{1i*} \vec{\mathbf{e}}_{1i*}$  on  $\psi$

$$\psi_{1i*} = \lambda_{\max_\psi}^{-1} \text{comp}_{\mathbf{x}_{1i*}} \left( \overrightarrow{\tilde{\psi}_{1i*} \|\tilde{\mathbf{x}}_*\|_{1i*}} \right) \|\mathbf{x}_{1i*}\|,$$

is shaped by an eigen-balanced first and second order statistical moment about the geometric locus of the correlated extreme vector  $\mathbf{x}_{1i*}$ .

Now, take any given Wolfe dual  $\psi_{1i*} \vec{\mathbf{e}}_{1i*}$  and correlated constrained primal  $\psi_{1i*} \mathbf{x}_{1i*}$  normal eigenaxis component. It will now be shown that the direction of  $\psi_{1i*} \vec{\mathbf{e}}_{1i*}$  is identical to the direction of  $\psi_{1i*} \mathbf{x}_{1i*}$ .

## 14.2 Directional Symmetries Exhibited by Normal Eigenaxis Components on $\psi$ and $\tau_1$

The vector direction of the Wolfe dual  $\psi_{1i*} \vec{\mathbf{e}}_{1i*}$  normal eigenaxis component is implicitly specified by Eq. (61), where it has been assumed that  $\psi_{1i*}$  provides an eigen-scaling for a non-orthogonal unit vector  $\vec{\mathbf{e}}_{1i*}$ . Given Eqs (51) and (59), it follows that the eigen-balanced pointwise covariance statistic in Eq. (61) encodes the direction of the extreme vector  $\mathbf{x}_{1i*}$  and an eigen-balanced magnitude along the axis of the extreme vector  $\mathbf{x}_{1i*}$ .

Returning to Eqs (64), (65), and (66), take any given Wolfe dual normal eigenaxis component  $\psi_{1i*} \vec{\mathbf{e}}_{1i*}$  that is correlated with an extreme vector  $\mathbf{x}_{1i*}$ . Given that the length  $\psi_{1i*}$  of the Wolfe dual normal eigenaxis component  $\psi_{1i*} \vec{\mathbf{e}}_{1i*}$  is determined by the eigen-balanced length of the extreme vector  $\mathbf{x}_{1i*}$

$$\psi_{1i*} = \lambda_{\max_\psi}^{-1} \text{comp}_{\mathbf{x}_{1i*}} \left( \overrightarrow{\tilde{\psi}_{1i*} \|\tilde{\mathbf{x}}_*\|_{1i*}} \right) \|\mathbf{x}_{1i*}\|,$$

it follows that the non-orthogonal unit vector  $\vec{\mathbf{e}}_{1i*}$  has the same direction as the extreme vector  $\mathbf{x}_{1i*}$

$$\vec{\mathbf{e}}_{1i*} \equiv \frac{\mathbf{x}_{1i*}}{\|\mathbf{x}_{1i*}\|}.$$

Thus, the direction of the Wolfe dual normal eigenaxis component  $\psi_{1i*} \vec{\mathbf{e}}_{1i*}$  is identical to the direction of the constrained primal normal eigenaxis component  $\psi_{1i*} \mathbf{x}_{1i*}$ , which is determined by the direction of the eigen-scaled  $\psi_{1i*}$  extreme training vector  $\mathbf{x}_{1i*}$ . The Wolfe dual  $\psi_{1i*} \vec{\mathbf{e}}_{1i*}$  and the constrained primal  $\psi_{1i*} \mathbf{x}_{1i*}$  normal eigenaxis components are said to exhibit directional

symmetry. Accordingly, each Wolfe dual  $\psi_{1i*} \vec{e}_{1i*}$  and correlated constrained primal  $\psi_{1i*} \mathbf{x}_{1i*}$  normal eigenaxis component exhibit directional symmetry.

It is concluded that the uniform directions of the Wolfe dual  $\psi_{1i*} \vec{e}_{1i*}$  and the constrained primal  $\psi_{1i*} \mathbf{x}_{1i*}$  normal eigenaxis components determine critical directions of large covariance, which contribute to a symmetric partitioning of a minimal geometric region of constant width that spans a region of large covariance between the distributions of two classes of training data. It is also concluded that each of the correlated normal eigenaxis components on  $\psi$  and  $\tau_1$  possess critical lengths for which the constrained discriminant function  $D(\mathbf{x}) = \mathbf{x}^T \tau + \tau_0$  delineates centrally located, bipartite, symmetric regions of large covariance between two data distributions. Expressions for the integrated lengths  $\sum_{i=1}^{l_1} \psi_{1i*}$  of the  $\psi_{1i*} \vec{e}_{1i*}$  components are obtained next.

### Integrated Lengths of $\psi_{1i*} \vec{e}_{1i*}$ Components on $\psi$

Using Eq. (61), an expression is obtained for the integrated lengths  $\sum_{i=1}^{l_1} \psi_{1i*}$  of the  $\psi_{1i*} \vec{e}_{1i*}$  Wolfe dual normal eigenaxis components on  $\psi$  :

$$\begin{aligned} \sum_{i=1}^{l_1} \psi_{1i*} &= \lambda_{\max \psi}^{-1} \sum_{i=1}^{l_1} \|\mathbf{x}_{1i*}\| \\ &\times \sum_{j=1}^{l_1} \psi_{1j*} \|\mathbf{x}_{1j*}\| \cos \theta_{\mathbf{x}_{1i*} \mathbf{x}_{1j*}} \\ &- \lambda_{\max \psi}^{-1} \sum_{i=1}^{l_1} \|\mathbf{x}_{1i*}\| \\ &\times \sum_{j=1}^{l_2} \psi_{2j*} \|\mathbf{x}_{2j*}\| \cos \theta_{\mathbf{x}_{1i*} \mathbf{x}_{2j*}}, \end{aligned} \quad (67)$$

where the direction of the Wolfe dual normal eigenaxis component  $\psi_{1i*} \vec{e}_{1i*}$  on  $\psi \in \mathbb{R}^N$  has the same direction as the correlated extreme training vector  $\mathbf{x}_{1i*} \in \mathbb{R}^d$ .

Alternatively, Eq. (63) provides the expression

$$\begin{aligned} \sum_{i=1}^{l_1} \psi_{1i*} &= \lambda_{\max \psi}^{-1} \sum_{i=1}^{l_1} \|\mathbf{x}_{1i*}\| \\ &\times \sum_{j=1}^{l_1} \cos \theta_{\mathbf{x}_{1i*} \mathbf{x}_{1j*}} \|\tau_1(j)\| \\ &- \lambda_{\max \psi}^{-1} \sum_{i=1}^{l_1} \|\mathbf{x}_{1i*}\| \\ &\times \sum_{j=1}^{l_2} \cos \theta_{\mathbf{x}_{1i*} \mathbf{x}_{2j*}} \|\tau_2(j)\|; \end{aligned} \quad (68)$$

Eqs (67) and (68) will be used to examine the algebraic and geometric nature of the statistical equilibrium state that is implied by Eq. (38). The uniform geometric and statistical properties which are jointly exhibited by the Wolfe dual  $\psi_{1i*} \vec{e}_{1i*}$  and the constrained primal  $\psi_{1i*} \mathbf{x}_{1i*}$  normal eigenaxis components are summarized below.

## Summary of Uniform Geometric and Statistical Properties Jointly Exhibited by Normal Eigenaxis Components on $\psi$ and $\tau_1$

Results of the previous analysis are now used to identify uniform geometric and statistical properties which are jointly exhibited by the  $\psi_{1i*} \vec{\mathbf{e}}_{1i*}$  Wolfe dual normal eigenaxis components on  $\psi$  and the  $\psi_{1i*} \mathbf{x}_{1i*}$  constrained primal normal eigenaxis components on  $\tau_1$ . The properties are summarized below.

**Conclusion 1** *The direction of each Wolfe dual normal eigenaxis component  $\psi_{1i*} \vec{\mathbf{e}}_{1i*}$  on  $\psi \in \mathbb{R}^N$  is identical to the direction of a constrained primal normal eigenaxis component  $\psi_{1i*} \mathbf{x}_{1i*}$  on  $\tau_1 \in \mathbb{R}^d$ .*

**Conclusion 2** *The lengths of each Wolfe dual normal eigenaxis component  $\psi_{1i*} \vec{\mathbf{e}}_{1i*}$  on  $\psi \in \mathbb{R}^N$  and correlated constrained primal normal eigenaxis component  $\psi_{1i*} \mathbf{x}_{1i*}$  on  $\tau_1 \in \mathbb{R}^d$  are shaped by identical joint symmetrical distributions of normal eigenaxis components on  $\psi$  and  $\tau$ .*

**Conclusion 3** *The length  $\psi_{1i*}$  of each Wolfe dual normal eigenaxis component  $\psi_{1i*} \vec{\mathbf{e}}_{1i*}$  on  $\psi \in \mathbb{R}^N$*

$$\psi_{1i*} = \lambda_{\max_\psi}^{-1} \text{comp}_{\vec{\mathbf{x}}_{1i*}} \left( \overrightarrow{\tilde{\psi}_{1i*} \|\tilde{\mathbf{x}}_*\|_{1i*}} \right) \|\mathbf{x}_{1i*}\|,$$

*is shaped by an eigen-balanced pointwise covariance estimate*

$$\begin{aligned} \widehat{\text{cov}}_{up\downarrow}(\mathbf{x}_{1i*}) &= \lambda_{\max_\psi}^{-1} \|\mathbf{x}_{1i*}\| \\ &\times \sum_{j=1}^{l_1} \psi_{1j*} \|\mathbf{x}_{1j*}\| \cos \theta_{\mathbf{x}_{1i*} \mathbf{x}_{1j*}} \\ &- \lambda_{\max_\psi}^{-1} \|\mathbf{x}_{1i*}\| \\ &\times \sum_{j=1}^{l_2} \psi_{2j*} \|\mathbf{x}_{2j*}\| \cos \theta_{\mathbf{x}_{1i*} \mathbf{x}_{2j*}}, \end{aligned}$$

*for a correlated extreme training vector  $\mathbf{x}_{1i*} \in \mathbb{R}^d$ , such that the eigenlocus of each constrained primal normal eigenaxis component  $\psi_{1i*} \mathbf{x}_{1i*}$  on  $\tau_1 \in \mathbb{R}^d$  provides a maximum covariance estimate in a principal location, in the form of an eigen-balanced first and second order statistical moment about the geometric locus of an extreme data point  $\mathbf{x}_{1i*}$ .*

**Conclusion 4** *Each Wolfe dual normal eigenaxis component  $\psi_{1i*} \vec{\mathbf{e}}_{1i*}$  on  $\psi$  encodes an eigen-balanced first and second order statistical moment about the locus of a correlated extreme data point  $\mathbf{x}_{1i*}$ , relative to the eigenloci of all of the eigen-scaled extreme training points, which determines the eigenlocus of a constrained primal normal eigenaxis component  $\psi_{1i*} \mathbf{x}_{1i*}$  on  $\tau_1$ .*

**Conclusion 5** *Any given eigen-balanced pointwise covariance estimate  $\widehat{\text{cov}}_{up\downarrow}(\mathbf{x}_{1i*})$  encodes a distribution of first order coordinates for an extreme training vector  $\mathbf{x}_{1i*}$ , relative to the eigen-scaled extreme training vectors for a given data set. The distribution of first order coordinates for  $\mathbf{x}_{1i*}$  describes how the components of  $\mathbf{x}_{1i*}$  are distributed within the given collection of eigen-scaled extreme vectors.*

**Conclusion 6** *Returning to Figs 12 and 23, the integrated eigenloci of the  $\psi_{1i*} \vec{e}_{1i*}$  Wolfe dual normal eigenaxis components jointly and implicitly specify the geometric locus of a hyperplane decision border  $H_{+1}$ . Likewise, the integrated eigenloci of the  $\psi_{1i*} \vec{e}_{1i*}$  Wolfe dual normal eigenaxis components jointly and implicitly account for a symmetric partitioning of a minimal area surface of large covariance that is delineated by a hyperplane decision boundary  $H_0$  which is symmetrically located between a pair of hyperplane decision borders  $H_{+1}$  and  $H_{-1}$ .*

**Claim 7** *The square  $\|\psi_{1i*} \mathbf{x}_{1i*}\|_{\min_c}^2$  of the constrained primal normal eigenaxis component  $\psi_{1i*} \mathbf{x}_{1i*}$  on  $\tau_1 \in \mathbb{R}^d$  is the probability of finding the extreme data point  $\mathbf{x}_{1i*}$  in a particular region of  $\mathbb{R}^d$ , where  $\|\psi_{1i*} \mathbf{x}_{1i*}\|_{\min_c}^2$  is the total allowed eigenenergy of  $\psi_{1i*} \mathbf{x}_{1i*}$ .*

The geometric and statistical properties of the eigenloci of the  $\psi_{2i*} \vec{e}_{2i*}$  Wolfe dual normal eigenaxis components on  $\psi$  are examined in the next section.

## 15 Eigenloci of the $\psi_{2i*} \vec{e}_{2i*}$ Wolfe Dual Normal Eigenaxis Components

Let  $i = 1 : l_2$ , where the extreme pattern vector  $\mathbf{x}_{2i*}$  has the training label  $y_i = -1$ . Using Eqs (57) and (58), the eigenlocus of the  $i^{th}$  Wolfe dual normal eigenaxis component  $\psi_{2i*} \vec{e}_{2i*}$  on  $\psi$  is a function of the expression:

$$\begin{aligned} \psi_{2i*} &= \lambda_{\max_\psi}^{-1} \|\mathbf{x}_{2i*}\| \sum_{j=1}^{l_2} \psi_{2j*} \|\mathbf{x}_{2j*}\| \cos \theta_{\mathbf{x}_{2i*} \mathbf{x}_{2j*}} \\ &\quad - \lambda_{\max_\psi}^{-1} \|\mathbf{x}_{2i*}\| \sum_{j=1}^{l_1} \psi_{1j*} \|\mathbf{x}_{1j*}\| \cos \theta_{\mathbf{x}_{2i*} \mathbf{x}_{1j*}}, \end{aligned} \quad (69)$$

where  $\psi_{2i*}$  provides an eigen-scaling for the non-orthogonal unit vector  $\vec{e}_{2i*}$ . Geometric and statistical explanations for the eigenlocus statistics:

$$\psi_{2j*} \|\mathbf{x}_{2j*}\| \cos \theta_{\mathbf{x}_{2i*} \mathbf{x}_{2j*}} \text{ and } \psi_{1j*} \|\mathbf{x}_{1j*}\| \cos \theta_{\mathbf{x}_{2i*} \mathbf{x}_{1j*}} \quad (70)$$

in Eq. (69) are considered next.

### Geometric and Statistical Explanations of $\psi_{2i*} \vec{e}_{2i*}$ Eigenlocus Statistics

The first geometric interpretation of the eigenlocus statistics in Eq. (70) defines the terms:

$$\psi_{2j*} \text{ and } \psi_{1j*},$$

to be eigen-scales for the signed magnitudes of the vector projections:

$$\|\mathbf{x}_{2j*}\| \cos \theta_{\mathbf{x}_{2i*} \mathbf{x}_{2j*}} \text{ and } \|\mathbf{x}_{1j*}\| \cos \theta_{\mathbf{x}_{2i*} \mathbf{x}_{1j*}},$$

of the eigen-scaled extreme vectors  $\psi_{2j*} \mathbf{x}_{2j*}$  and  $\psi_{1j*} \mathbf{x}_{1j*}$  along the axis of the extreme vector  $\mathbf{x}_{2i*}$ , where  $\cos \theta_{\mathbf{x}_{2i*} \mathbf{x}_{2j*}}$  and  $\cos \theta_{\mathbf{x}_{2i*} \mathbf{x}_{1j*}}$  are the angles between the axes of the eigen-scaled extreme vectors  $\psi_{2j*} \mathbf{x}_{2j*}$  and  $\psi_{1j*} \mathbf{x}_{1j*}$  and the axis of the extreme vector  $\mathbf{x}_{2i*}$ .

### An Alternative Geometric Interpretation

An alternative geometric explanation for the eigenlocus statistics in Eq. (70) accounts for the representation of the  $\tau_1$  and  $\tau_2$  primal normal eigenlocus components within the Wolfe dual eigenspace. Consider the algebraic relationships

$$\psi_{1j*} \|\mathbf{x}_{1j*}\| = \|\psi_{1j*} \mathbf{x}_{1j*}\| = \|\tau_1(j)\|,$$

and

$$\psi_{2j*} \|\mathbf{x}_{2j*}\| = \|\psi_{2j*} \mathbf{x}_{2j*}\| = \|\tau_2(j)\|,$$

where  $\tau_1(j)$  and  $\tau_2(j)$  are the  $j$ th constrained primal normal eigenaxis components on  $\tau_1$  and  $\tau_2$ . Given the above expressions, it follows that the eigen-scaled  $\psi_{2j*}$  signed magnitude  $\|\mathbf{x}_{2j*}\| \cos \theta_{\mathbf{x}_{2i*} \mathbf{x}_{2j*}}$  of the vector projection of the eigen-scaled extreme vector  $\psi_{2j*} \mathbf{x}_{2j*}$  along the axis of the extreme vector  $\mathbf{x}_{2i*}$

$$\psi_{2j*} \|\mathbf{x}_{2j*}\| \cos \theta_{\mathbf{x}_{2i*} \mathbf{x}_{2j*}},$$

encodes the cosine-scaled  $\cos \theta_{\mathbf{x}_{2i*} \mathbf{x}_{2j*}}$  length of the  $j$ th constrained primal normal eigenaxis component  $\tau_2(j)$  on  $\tau_2$

$$\cos \theta_{\mathbf{x}_{2i*} \mathbf{x}_{2j*}} \|\tau_2(j)\|,$$

where  $\psi_{2j*}$  is the magnitude of the  $\psi_{2j*} \vec{\mathbf{e}}_{2j*}$  Wolfe dual normal eigenaxis component, and  $\cos \theta_{\mathbf{x}_{2i*} \mathbf{x}_{2j*}}$  encodes the angle between the extreme training vectors  $\mathbf{x}_{2i*}$  and  $\mathbf{x}_{2j*}$ . Likewise, the eigen-scaled  $\psi_{1j*}$  signed magnitude  $\|\mathbf{x}_{1j*}\| \cos \theta_{\mathbf{x}_{2i*} \mathbf{x}_{1j*}}$  of the vector projection of the the eigen-scaled extreme vector  $\psi_{1j*} \mathbf{x}_{1j*}$  along the axis of the extreme vector  $\mathbf{x}_{2i*}$

$$\psi_{1j*} \|\mathbf{x}_{1j*}\| \cos \theta_{\mathbf{x}_{2i*} \mathbf{x}_{1j*}},$$

encodes the cosine-scaled  $\cos \theta_{\mathbf{x}_{2i*} \mathbf{x}_{1j*}}$  length of the  $j$ th constrained primal normal eigenaxis component  $\tau_1(j)$  on  $\tau_1$

$$\cos \theta_{\mathbf{x}_{2i*} \mathbf{x}_{1j*}} \|\tau_1(j)\|,$$

where  $\psi_{1j*}$  is the magnitude of the  $\psi_{1j*} \vec{\mathbf{e}}_{1j*}$  Wolfe dual normal eigenaxis component, and  $\cos \theta_{\mathbf{x}_{2i*} \mathbf{x}_{1j*}}$  encodes the angle between the extreme training vectors  $\mathbf{x}_{2i*}$  and  $\mathbf{x}_{1j*}$ .

It follows that the eigenlocus of the Wolfe dual normal eigenaxis component  $\psi_{2i*} \vec{\mathbf{e}}_{2i*}$  is a function of the constrained primal normal eigenaxis components

on  $\tau_1$  and  $\tau_2$ :

$$\begin{aligned}\psi_{2i*} &= \lambda_{\max\psi}^{-1} \|\mathbf{x}_{2i*}\| \sum_{j=1}^{l_2} \cos \theta_{\mathbf{x}_{2i*} \mathbf{x}_{2j*}} \|\tau_2(j)\| \\ &\quad - \lambda_{\max\psi}^{-1} \|\mathbf{x}_{2i*}\| \sum_{j=1}^{l_1} \cos \theta_{\mathbf{x}_{2i*} \mathbf{x}_{1j*}} \|\tau_1(j)\|.\end{aligned}\tag{71}$$

The next analysis considers how uniform geometric and statistical properties which are jointly exhibited by the Wolfe dual  $\psi_{2i*} \vec{\mathbf{e}}_{2i*}$  and the constrained primal  $\psi_{2i*} \mathbf{x}_{2i*}$  normal eigenaxis components account for a bipartite symmetric partitioning of a region of large covariance that is well-positioned between a pair of data distributions.

### 15.1 Uniform Geometric and Statistical Properties Jointly Exhibited by Normal Eigenaxis Components on $\psi$ and $\tau_2$

Using results from the previous analysis, the KKT constraint  $\sum_{i=1}^{l_1} \psi_{1i*} = \sum_{i=1}^{l_2} \psi_{2i*}$  of Eq. (38) indicates that Eq. (69) determines an eigen-balanced, signed magnitude along the axis of an extreme vector  $\mathbf{x}_{2i*}$ .

Let  $\text{comp}_{\mathbf{x}_{2i*}} \left( \overrightarrow{\psi_{2i*} \|\tilde{\mathbf{x}}_*\|_{2i*}} \right)$  denote the eigen-balanced, signed magnitude

$$\begin{aligned}\text{comp}_{\mathbf{x}_{2i*}} \left( \overrightarrow{\psi_{2i*} \|\tilde{\mathbf{x}}_*\|_{2i*}} \right) &= \sum_{j=1}^{l_2} \psi_{2j*} \\ &\quad \times \left[ \|\mathbf{x}_{2j*}\| \cos \theta_{\mathbf{x}_{2i*} \mathbf{x}_{2j*}} \right] \\ &\quad - \sum_{j=1}^{l_1} \psi_{1j*} \\ &\quad \times \left[ \|\mathbf{x}_{1j*}\| \cos \theta_{\mathbf{x}_{2i*} \mathbf{x}_{1j*}} \right],\end{aligned}\tag{72}$$

along the axis of the extreme training vector  $\mathbf{x}_{2i*}$  that is correlated with the Wolfe dual normal eigenaxis component  $\psi_{2i*} \vec{\mathbf{e}}_{2i*}$ .

Given Eqs (38) and (60), it follows that Eq. (72) determines an eigen-balanced distribution of eigen-scaled first degree coordinates of extreme training vectors along the axis of  $\mathbf{x}_{2i*}$ ; Eqs (38) and (60) also indicate that Eq. (72) determines an eigen-balanced first and second order statistical moment about the geometric locus of  $\mathbf{x}_{2i*}$ .

Furthermore, Eqs (38), (69), and (72) show that joint distributions of the components of  $\psi$  and  $\tau$  are symmetrically distributed over the axis of the extreme vector  $\mathbf{x}_{2i*}$ . This indicates that joint distributions of the components of  $\psi$  and  $\tau$  are symmetrically distributed over the axis of the Wolfe dual normal eigenaxis component  $\psi_{2i*} \vec{\mathbf{e}}_{2i*}$ .

Alternatively, returning to Eq. (71), the eigen-balanced, signed magnitude in Eq. (72) depends upon the difference between integrated, cosine-scaled mag-



nitudes of the constrained primal normal eigenaxis components on  $\tau_2$  and  $\tau_1$ :

$$\begin{aligned} \text{comp}_{\mathbf{x}_{2i*}} \left( \overrightarrow{\tilde{\psi}_{2i*} \|\tilde{\mathbf{x}}_*\|_{2i*}} \right) &= \sum_{j=1}^{l_2} \cos \theta_{\mathbf{x}_{2i*} \mathbf{x}_{2j*}} \|\tau_2(j)\| \\ &\quad - \sum_{j=1}^{l_1} \cos \theta_{\mathbf{x}_{2i*} \mathbf{x}_{1j*}} \|\tau_1(j)\|, \end{aligned} \quad (73)$$

which also shows that joint distributions of the components of  $\psi$  and  $\tau$  are symmetrically distributed over the axes of both  $\psi_{2i*} \vec{\mathbf{e}}_{2i*}$  and  $\mathbf{x}_{2i*}$ .

Using Eqs (69) and (72), it follows that the length  $\psi_{2i*}$  of the Wolfe dual normal eigenaxis component  $\psi_{2i*} \vec{\mathbf{e}}_{2i*}$  is determined by the weighted length of the correlated extreme training vector  $\mathbf{x}_{2i*}$

$$\psi_{2i*} = \left[ \lambda_{\max_\psi}^{-1} \times \text{comp}_{\mathbf{x}_{2i*}} \left( \overrightarrow{\tilde{\psi}_{2i*} \|\tilde{\mathbf{x}}_*\|_{2i*}} \right) \right] \|\mathbf{x}_{2i*}\|, \quad (74)$$

where the weighting factor encodes an eigenvalue  $\lambda_{\max_\psi}^{-1}$  scaling of an eigenbalanced, signed magnitude  $\text{comp}_{\mathbf{x}_{2i*}} \left( \overrightarrow{\tilde{\psi}_{2i*} \|\tilde{\mathbf{x}}_*\|_{2i*}} \right)$  along the axis of  $\mathbf{x}_{2i*}$ .

Given that  $\psi_{2i*} > 0$ ,  $\lambda_{\max_\psi}^{-1} > 0$ , and  $\|\mathbf{x}_{2i*}\| > 0$ , it follows that the eigenbalanced, signed magnitude along the axis of  $\mathbf{x}_{2i*}$  is positive

$$\text{comp}_{\mathbf{x}_{2i*}} \left( \overrightarrow{\tilde{\psi}_{2i*} \|\tilde{\mathbf{x}}_*\|_{2i*}} \right) > 0,$$

which indicates that the weighting factor in Eq. (74) determines an eigenbalanced length

$$\lambda_{\max_\psi}^{-1} \text{comp}_{\mathbf{x}_{2i*}} \left( \overrightarrow{\tilde{\psi}_{2i*} \|\tilde{\mathbf{x}}_*\|_{2i*}} \right) \|\mathbf{x}_{2i*}\|,$$

for the extreme vector  $\mathbf{x}_{2i*}$ . It follows that Eq. (74) determines an eigenbalanced length for  $\psi_{2i*} \vec{\mathbf{e}}_{2i*}$  and  $\mathbf{x}_{2i*}$ .

Returning to Eqs (60) and (69), it follows that the length  $\psi_{2i*}$  of the Wolfe dual normal eigenaxis component  $\psi_{2i*} \vec{\mathbf{e}}_{2i*}$  on  $\psi$

$$\psi_{2i*} = \lambda_{\max_\psi}^{-1} \text{comp}_{\mathbf{x}_{2i*}} \left( \overrightarrow{\tilde{\psi}_{2i*} \|\tilde{\mathbf{x}}_*\|_{2i*}} \right) \|\mathbf{x}_{2i*}\|,$$

is shaped by an eigen-balanced first and second order statistical moment about the geometric locus of the correlated extreme vector  $\mathbf{x}_{2i*}$ .

Now, take any given Wolfe dual  $\psi_{2i*} \vec{\mathbf{e}}_{2i*}$  and correlated constrained primal  $\psi_{2i*} \mathbf{x}_{2i*}$  normal eigenaxis component. It will now be shown that the direction of  $\psi_{2i*} \vec{\mathbf{e}}_{2i*}$  is identical to the direction of  $\psi_{2i*} \mathbf{x}_{2i*}$ .

## 15.2 Directional Symmetries Exhibited by Normal Eigenaxis Components on $\psi$ and $\tau_2$

The vector direction of the  $\psi_{2i*} \vec{\mathbf{e}}_{2i*}$  component is implicitly specified by Eq. (69), where it has been assumed that  $\psi_{2i*}$  provides an eigen-scale for a non-

orthogonal unit vector  $\vec{\mathbf{e}}_{2i*}$ . Given Eqs (51) and (60), it follows that the eigen-balanced pointwise covariance statistic in Eq. (69) encodes the direction of the extreme vector  $\mathbf{x}_{2i*}$  and an eigen-balanced magnitude along the axis of the extreme vector  $\mathbf{x}_{2i*}$ .

Returning to Eqs (72), (73), and (74), take any given Wolfe dual normal eigenaxis component  $\psi_{2i*} \vec{\mathbf{e}}_{2i*}$  that is correlated with an extreme vector  $\mathbf{x}_{2i*}$ . Given that the length  $\psi_{2i*}$  of the Wolfe dual normal eigenaxis component  $\psi_{2i*} \vec{\mathbf{e}}_{2i*}$  is determined by the eigen-balanced length of the extreme vector  $\mathbf{x}_{2i*}$

$$\psi_{2i*} = \lambda_{\max_\psi}^{-1} \text{comp}_{\vec{\mathbf{x}}_{2i*}} \left( \overrightarrow{\psi_{2*} \|\tilde{\mathbf{x}}_*\|_{2*}} \right) \|\mathbf{x}_{2i*}\|,$$

it follows that the non-orthogonal unit vector  $\vec{\mathbf{e}}_{2i*}$  has the same direction as the extreme vector  $\mathbf{x}_{2i*}$

$$\vec{\mathbf{e}}_{2i*} \equiv \frac{\mathbf{x}_{2i*}}{\|\mathbf{x}_{2i*}\|}.$$

Therefore, the direction of the Wolfe dual normal eigenaxis component  $\psi_{2i*} \vec{\mathbf{e}}_{2i*}$  is identical to the direction of the constrained primal normal eigenaxis component  $\psi_{2i*} \mathbf{x}_{2i*}$ , which is determined by the direction of the eigen-scaled  $\psi_{2i*}$  extreme training vector  $\mathbf{x}_{2i*}$ . The Wolfe dual  $\psi_{2i*} \vec{\mathbf{e}}_{1i*}$  and the constrained primal  $\psi_{2i*} \mathbf{x}_{2i*}$  normal eigenaxis components are said to exhibit directional symmetry. Accordingly, each Wolfe dual  $\psi_{2i*} \vec{\mathbf{e}}_{2i*}$  and correlated constrained primal  $\psi_{2i*} \mathbf{x}_{2i*}$  normal eigenaxis component exhibit directional symmetry.

It is concluded that the uniform directions of the Wolfe dual  $\psi_{2i*} \vec{\mathbf{e}}_{1i*}$  and the constrained primal  $\psi_{2i*} \mathbf{x}_{2i*}$  normal eigenaxis components determine critical directions of large covariance, which contribute to a symmetric partitioning of a minimal geometric region of constant width that spans a region of large covariance between the distributions of two classes of training data. It is also concluded that each of the correlated normal eigenaxis components on  $\psi$  and  $\tau_2$  possess critical lengths for which the constrained discriminant function  $D(\mathbf{x}) = \mathbf{x}^T \tau + \tau_0$  delineates centrally located, bipartite, symmetric regions of large covariance between two data distributions. Equations for the integrated lengths  $\sum_{i=1}^{l_2} \psi_{2i*}$  of the  $\psi_{2i*} \vec{\mathbf{e}}_{2i*}$  components are obtained next.

### Integrated Lengths of $\psi_{2i*} \vec{\mathbf{e}}_{2i*}$ Components on $\psi$

Using Eq. (69), an equation is obtained for the integrated lengths  $\sum_{i=l_1+1}^{l_2} \psi_{2i*}$  of the  $\psi_{2i*} \vec{\mathbf{e}}_{2i*}$  components:

$$\begin{aligned} \sum_{i=1}^{l_2} \psi_{2i*} &= \lambda_{\max_\psi}^{-1} \sum_{i=1}^{l_2} \|\mathbf{x}_{2i*}\| \\ &\times \sum_{j=1}^{l_2} \psi_{2j*} \|\mathbf{x}_{2j*}\| \cos \theta_{\mathbf{x}_{2i*} \mathbf{x}_{2j*}} \\ &- \lambda_{\max_\psi}^{-1} \sum_{i=1}^{l_2} \|\mathbf{x}_{2i*}\| \\ &\times \sum_{j=1}^{l_1} \psi_{1j*} \|\mathbf{x}_{1j*}\| \cos \theta_{\mathbf{x}_{2i*} \mathbf{x}_{1j*}}, \end{aligned} \quad (75)$$

where the vector direction of the Wolfe dual normal eigenaxis  $\psi_{2i*} \vec{e}_{1i*}$  on  $\psi \in \mathbb{R}^N$  has the same direction as the correlated extreme training vector  $\mathbf{x}_{2i*} \in \mathbb{R}^d$ .

Alternatively, Eq. (71) provides the expression:

$$\begin{aligned} \sum_{i=1}^{l_2} \psi_{2i*} &= \lambda_{\max_\psi}^{-1} \sum_{i=1}^{l_2} \|\mathbf{x}_{2i*}\| \\ &\times \sum_{j=1}^{l_2} \cos \theta_{\mathbf{x}_{2i*} \mathbf{x}_{2j*}} \|\tau_2(j)\| \\ &- \lambda_{\max_\psi}^{-1} \sum_{i=1}^{l_2} \|\mathbf{x}_{2i*}\| \\ &\times \sum_{j=1}^{l_1} \cos \theta_{\mathbf{x}_{2i*} \mathbf{x}_{1j*}} \|\tau_1(j)\|; \end{aligned} \quad (76)$$

Eqs (75) and (76) will be used to examine the algebraic and geometric nature of the statistical equilibrium state that is implied by Eq. (38). The uniform geometric and statistical properties which are jointly exhibited by the  $\psi_{2i*} \vec{e}_{2i*}$  and the  $\psi_{2i*} \mathbf{x}_{2i*}$  normal eigenaxis components are summarized next.

### Summary of Uniform Geometric and Statistical Properties Jointly Exhibited by Normal Eigenaxis Components on $\psi$ and $\tau_2$

Results of the previous analysis are now used to identify uniform geometric and statistical properties which are jointly exhibited by the  $\psi_{2i*} \vec{e}_{2i*}$  Wolfe dual normal eigenaxis components on  $\psi$  and the  $\psi_{2i*} \mathbf{x}_{2i*}$  constrained primal normal eigenaxis components on  $\tau_2$ . The properties are summarized below.

**Conclusion 8** *The direction of each Wolfe dual normal eigenaxis component  $\psi_{2i*} \vec{e}_{2i*}$  on  $\psi \in \mathbb{R}^N$  is identical to the direction of a constrained primal normal eigenaxis component  $\psi_{2i*} \mathbf{x}_{2i*}$  on  $\tau_2 \in \mathbb{R}^d$ .*

**Conclusion 9** *The lengths of each Wolfe dual normal eigenaxis component  $\psi_{2i*} \vec{e}_{2i*}$  on  $\psi \in \mathbb{R}^N$  and correlated constrained primal normal eigenaxis component  $\psi_{2i*} \mathbf{x}_{2i*}$  on  $\tau_2 \in \mathbb{R}^d$  are shaped by identical joint symmetrical distributions of normal eigenaxis components on  $\psi$  and  $\tau$ .*

**Conclusion 10** *The length  $\psi_{2i*}$  of each Wolfe dual normal eigenaxis component  $\psi_{2i*} \vec{e}_{2i*}$  on  $\psi \in \mathbb{R}^N$*

$$\psi_{2i*} = \lambda_{\max_\psi}^{-1} \text{comp}_{\vec{\mathbf{x}}_{2i*}} \left( \overrightarrow{\tilde{\psi}_{2i*} \|\tilde{\mathbf{x}}_*\|_{2i*}} \right) \|\mathbf{x}_{2i*}\|,$$

*is shaped by an eigen-balanced pointwise covariance estimate*

$$\begin{aligned} \widehat{\text{cov}}_{up\uparrow}(\mathbf{x}_{2i*}) &= \lambda_{\max_\psi}^{-1} \|\mathbf{x}_{2i*}\| \\ &\times \sum_{j=1}^{l_2} \psi_{2j*} \|\mathbf{x}_{2j*}\| \cos \theta_{\mathbf{x}_{2i*} \mathbf{x}_{2j*}} \\ &- \lambda_{\max_\psi}^{-1} \|\mathbf{x}_{2i*}\| \\ &\times \sum_{j=1}^{l_1} \psi_{1j*} \|\mathbf{x}_{1j*}\| \cos \theta_{\mathbf{x}_{2i*} \mathbf{x}_{1j*}}, \end{aligned}$$

for a correlated extreme training vector  $\mathbf{x}_{2_{i*}} \in \mathbb{R}^d$ , such that the eigenlocus of each constrained primal normal eigenaxis component  $\psi_{2_{i*}} \mathbf{x}_{2_{i*}}$  on  $\tau_2 \in \mathbb{R}^d$  provides a maximum covariance estimate in a principal location, in the form of an eigen-balanced first and second order statistical moment about the geometric locus of an extreme point  $\mathbf{x}_{2_{i*}}$ .

**Conclusion 11** Each Wolfe dual normal eigenaxis component  $\psi_{2_{i*}} \vec{\mathbf{e}}_{2_{i*}}$  on  $\psi$  encodes an eigen-balanced first and second order statistical moment about the geometric locus of a correlated extreme data point  $\mathbf{x}_{2_{i*}}$ , relative to the eigenloci of all of the eigen-scaled extreme training points, which determines the eigenlocus of a constrained primal normal eigenaxis component  $\psi_{2_{i*}} \mathbf{x}_{2_{i*}}$  on  $\tau_2$ .

**Conclusion 12** Any given eigen-balanced pointwise covariance estimate  $\widehat{\text{cov}}_{up\downarrow}(\mathbf{x}_{2_{i*}})$  encodes a distribution of first order coordinates for an extreme training vector  $\mathbf{x}_{2_{i*}}$ , relative to the eigen-scaled extreme training points for a given data set. The distribution of first order coordinates for  $\mathbf{x}_{2_{i*}}$  describes how the components of  $\mathbf{x}_{2_{i*}}$  are distributed within the given collection of eigen-scaled extreme vectors.

**Conclusion 13** Returning to Figs 12 and 23, the integrated eigenloci of the  $\psi_{2_{i*}} \vec{\mathbf{e}}_{2_{i*}}$  Wolfe dual normal eigenaxis components jointly and implicitly specify the geometric locus of a hyperplane decision border  $H_{-1}$ . Likewise, the integrated eigenloci of the  $\psi_{2_{i*}} \vec{\mathbf{e}}_{2_{i*}}$  Wolfe dual normal eigenaxis components jointly and implicitly account for a symmetric partitioning of a minimal area surface of large covariance that is delineated by a hyperplane decision boundary  $H_0$ , which is symmetrically located between a pair of hyperplane decision borders  $H_{+1}$  and  $H_{-1}$ .

**Claim 14** The square  $\|\psi_{2_{i*}} \mathbf{x}_{2_{i*}}\|_{\min_c}^2$  of the constrained primal normal eigenaxis component  $\psi_{2_{i*}} \mathbf{x}_{2_{i*}}$  on  $\tau_2 \in \mathbb{R}^d$  is the probability of finding the extreme data point  $\mathbf{x}_{2_{i*}}$  in a particular region of  $\mathbb{R}^d$ , where  $\|\psi_{2_{i*}} \mathbf{x}_{2_{i*}}\|_{\min_c}^2$  is the total allowed eigenenergy of  $\psi_{2_{i*}} \mathbf{x}_{2_{i*}}$ .

The properties exhibited by the total allowed eigenenergy of a Wolf dual normal eigenlocus are summarized in the next section. Section 16 will also outline the fundamental issue that must be resolved for strong dual normal eigenlocus transforms.

## 16 Properties Exhibited by the Total Allowed Eigenenergy of a Wolfe Dual Normal Eigenlocus $\psi$

The eigenloci of the Wolfe dual normal eigenaxis components  $\{\psi_{i*} \vec{\mathbf{e}}_{i*}\}_{i=1}^l$  on  $\psi$  determine the total allowed eigenenergy  $\|\psi\|_{\min_c}^2$  of  $\psi$

$$\begin{aligned}\|\psi\|_{\min_c}^2 &= \left( \sum_{i=1}^l \psi_{i*} \vec{\mathbf{e}}_{i*}^T \right) \left( \sum_{i=1}^l \psi_{i*} \vec{\mathbf{e}}_{i*} \right), \\ &= \sum_{i=1}^l \psi_{i*}^2.\end{aligned}$$

Given the uniform geometric and statistical properties which are jointly exhibited by correlated normal eigenaxis components on  $\tau_1 - \tau_2$  and  $\psi$ , it is concluded that *a Wolfe dual normal eigenlocus  $\psi$  satisfies a critical minimum eigenenergy constraint which is symmetrically related to the restriction of the primal normal eigenlocus to the Wolfe dual eigenspace.*

Therefore, consider again Eq. (54)

$$\begin{aligned}\max \psi^T \mathbf{Q} \psi &= \lambda_{\max_\psi} \|\psi\|_{\min_c}^2, \\ &\cong \|\tau\|_{\min_c}^2,\end{aligned}$$

and Eq. (38)

$$\sum_{i=1}^{l_1} \psi_{1i*} = \sum_{i=1}^{l_2} \psi_{2i*}.$$

Equations (38), (61), (64), (69), and (72), which demonstrate how joint distributions of the normal eigenaxis components of  $\psi$  and  $\tau$  are symmetrically distributed over the axes of all of the Wolfe dual normal eigenaxis components  $\{\psi_{1i*} \vec{\mathbf{e}}_{1i*}\}_{i=1}^{l_1}$  and  $\{\psi_{2i*} \vec{\mathbf{e}}_{2i*}\}_{i=1}^{l_2}$  and correlated extreme training vectors  $\{\mathbf{x}_{1i*}\}_{i=1}^{l_1}$  and  $\{\mathbf{x}_{2i*}\}_{i=1}^{l_2}$ , together with Eq. (54), indicate that the critical minimum eigenenergy  $\lambda_{\max_\psi} \|\psi\|_{\min_c}^2$  of  $\psi$  is characterized by joint symmetrical distributions of the eigenenergies of  $\psi$  and  $\tau$ , which are symmetrically distributed over the Wolfe dual normal eigenaxis components.

It follows that joint distributions of the normal eigenaxis components of  $\psi$  and  $\tau$  are symmetrically distributed over the axes of all of the constrained primal normal eigenaxis components  $\{\psi_{1i*} \mathbf{x}_{1i*}\}_{i=1}^{l_1}$  and  $\{\psi_{2i*} \mathbf{x}_{2i*}\}_{i=1}^{l_2}$ , whereby the critical minimum eigenenergy  $\|\tau\|_{\min_c}^2$  of the constrained primal normal eigenlocus  $\tau$  is also characterized by joint symmetrical distributions of the eigenenergies of  $\psi$  and  $\tau$ .

Section 17 will demonstrate how the total allowed eigenenergy  $\lambda_{\max_\psi} \|\psi\|_{\min_c}^2$  of  $\psi$ , which is determined by the eigenloci of the Wolfe dual normal eigenaxis components, regulates the manner in which joint symmetrical distributions of the eigenenergies of  $\psi$  and  $\tau$  are symmetrically distributed over the constrained primal normal eigenlocus components  $\tau_1$  and  $\tau_2$  on  $\tau$ .

The next section will outline the fundamental issue that must be resolved for strong dual normal eigenlocus transforms. Finding the state of statistical equilibrium for the constrained discriminant function of a strong dual normal eigenlocus  $\tau$  requires finding the right mix of normal eigenaxis component lengths on  $\psi$  and  $\tau$ . The fundamental problem involves determining the lengths of the Wolfe dual normal eigenaxis components on  $\psi$ , which are the Lagrange multipliers  $\{\psi_i\}_{i=1}^N$  of Eqs (39) and (40).

## 16.1 Finding the Right Component Lengths

It has been demonstrated that the directions of the constrained primal and the Wolfe dual normal eigenaxis components are fixed, along with the angles between all of the extreme vectors. Moreover, the equilibrium constraint of Eq. (38) on the component lengths of  $\psi$

$$\sum_{i=1}^{l_1} \psi_{1i*} = \sum_{i=1}^{l_2} \psi_{2i*},$$

indicates that the RHS of Eq. (67)

$$\begin{aligned} \sum_{i=1}^{l_1} \psi_{1i*} &= \lambda_{\max\psi}^{-1} \sum_{i=1}^{l_1} \|\mathbf{x}_{1i*}\| \\ &\times \sum_{j=1}^{l_1} \psi_{1j*} \|\mathbf{x}_{1j*}\| \cos \theta_{\mathbf{x}_{1i*} \mathbf{x}_{1j*}} \\ &- \lambda_{\max\psi}^{-1} \sum_{i=1}^{l_1} \|\mathbf{x}_{1i*}\| \\ &\times \sum_{j=1}^{l_2} \psi_{2j*} \|\mathbf{x}_{2j*}\| \cos \theta_{\mathbf{x}_{1i*} \mathbf{x}_{2j*}}, \end{aligned}$$

must equal the RHS of Eq. (75)

$$\begin{aligned} \sum_{i=1}^{l_2} \psi_{2i*} &= \lambda_{\max\psi}^{-1} \sum_{i=1}^{l_2} \|\mathbf{x}_{2i*}\| \\ &\times \sum_{j=1}^{l_2} \psi_{2j*} \|\mathbf{x}_{2j*}\| \cos \theta_{\mathbf{x}_{2i*} \mathbf{x}_{2j*}} \\ &- \lambda_{\max\psi}^{-1} \sum_{i=1}^{l_2} \|\mathbf{x}_{2i*}\| \\ &\times \sum_{j=1}^{l_1} \psi_{1j*} \|\mathbf{x}_{1j*}\| \cos \theta_{\mathbf{x}_{2i*} \mathbf{x}_{1j*}}. \end{aligned}$$

Likewise, the RHS of Eq. (68)

$$\begin{aligned} \sum_{i=1}^{l_1} \psi_{1i*} &= \lambda_{\max\psi}^{-1} \sum_{i=1}^{l_1} \|\mathbf{x}_{1i*}\| \\ &\times \sum_{j=1}^{l_1} \cos \theta_{\mathbf{x}_{1i*} \mathbf{x}_{1j*}} \|\tau_1(j)\| \\ &- \lambda_{\max\psi}^{-1} \sum_{i=1}^{l_1} \|\mathbf{x}_{1i*}\| \\ &\times \sum_{j=1}^{l_2} \cos \theta_{\mathbf{x}_{1i*} \mathbf{x}_{2j*}} \|\tau_2(j)\|, \end{aligned}$$

must equal the RHS of Eq. (76)

$$\begin{aligned}
\sum_{i=1}^{l_2} \psi_{2i*} &= \lambda_{\max\psi}^{-1} \sum_{i=1}^{l_2} \|\mathbf{x}_{2i*}\| \\
&\times \sum_{j=1}^{l_2} \cos \theta_{\mathbf{x}_{2i*} \mathbf{x}_{2j*}} \|\tau_2(j)\| \\
&- \lambda_{\max\psi}^{-1} \sum_{i=1}^{l_2} \|\mathbf{x}_{2i*}\| \\
&\times \sum_{j=1}^{l_1} \cos \theta_{\mathbf{x}_{2i*} \mathbf{x}_{1j*}} \|\tau_1(j)\|.
\end{aligned}$$

## 16.2 Critical Length Constraints

The pair of balanced statistical eigenlocus equations outlined directly above indicate that *all of the magnitudes*

$$\{\psi_{1i*} | \psi_{1i*} > 0\}_{i=1}^{l_1},$$

and

$$\{\psi_{2i*} | \psi_{2i*} > 0\}_{i=1}^{l_2},$$

of the Wolfe dual normal eigenaxis components on  $\psi$  satisfy critical length constraints, such that the highly interconnected sets of inner product relationships amongst the Wolfe dual and the constrained primal normal eigenaxis components in Eqs (67) and (75), or Eqs (68) and (76), determine proper lengths  $\psi_{1i*}$  or  $\psi_{2i*}$  for each Wolfe dual normal eigenaxis component  $\psi_{1i*} \vec{\mathbf{e}}_{1i*}$  or  $\psi_{2i*} \vec{\mathbf{e}}_{2i*}$ , which effectively determine the proper length of a correlated, constrained primal normal eigenaxis component  $\psi_{1i*} \mathbf{x}_{1i*}$  or  $\psi_{2i*} \mathbf{x}_{2i*}$ .

Recall that the real unknowns associated with the inequality constrained optimization problem in Eq. (13) are the constrained eigen-coordinate locations of a normal eigenaxis  $\mathbf{v}$ , which are essentially determined by the magnitudes or lengths of the Wolfe dual normal eigenaxis components on  $\psi$ . It has been shown that each constrained primal normal eigenaxis component on  $\tau$  is formed by an eigen-scaled extreme training vector, where the eigen-scale of each extreme training vector is the magnitude of a correlated Wolfe dual normal eigenaxis component on  $\psi$ . It has also been demonstrated that each extreme training vector and correlated Wolfe dual normal eigenaxis component exhibit directional symmetry.

All of the previous analyses and simulation studies indicate that the magnitudes of each of the Wolfe dual normal eigenaxis components on  $\psi$  exhibit symmetric proportions which determine properly proportioned magnitudes for each of the constrained primal normal eigenaxis components on  $\tau$ . It is claimed that the magnitudes of the Wolfe dual normal eigenaxis components essentially determine the state of statistical equilibrium which is exhibited by a strong dual normal eigenlocus  $\tau = \tau_1 - \tau_2$ . In particular, it is claimed that the state of equilibrium

$$\sum_{i=1}^{l_1} \psi_{1i*} = \sum_{i=1}^{l_2} \psi_{2i*},$$

that is satisfied by the component lengths of  $\psi$ , which ensures that joint distributions of the components of  $\psi$  and  $\tau$  are symmetrically distributed over each of the eigen-scaled extreme training vectors on  $\tau_1$  and  $\tau_2$ , determines a point of statistical equilibrium

$$\tau = \tau_1 - \tau_2,$$

for which a strong dual normal eigenlocus  $\tau = \tau_1 - \tau_2$  exhibits a critical minimum eigenenergy that is characterized by joint symmetrical distributions of the eigenenergies of  $\psi$  and  $\tau$ .

The next section of the paper will examine the algebraic, geometric, and statistical nature of the remarkable statistical balancing feat that is routinely accomplished by solving the inequality constrained optimization problem in Eq. (13). Algebraic expressions will be developed for the total allowed eigenenergies of  $\tau_1$ ,  $\tau_2$ , and  $\tau$ . These expressions will be used to develop statistical expressions for the statistical fulcrum  $f_s$  and the symmetric equalizer statistic  $\nabla_{eq}$  in Eq. (34). All of these results will be used to develop the statistical machinery behind the point of statistical equilibrium which is determined by the constrained Lagrangian functional  $L_{\Psi(\tau)}$  of Eq. (15). In the final sections of the analysis, an algebraic expression called the *strong dual normal eigenlocus identity* will be developed, in which the total allowed eigenenergies of a strong dual normal eigenlocus satisfy the law of cosines in a surprisingly elegant and symmetrical manner.

## 17 An Elegant Statistical Balancing Feat

The strong duality relationships between the constrained primal normal eigenaxis components on  $\tau \in \mathbb{R}^d$  and the Wolfe dual normal eigenaxis components on  $\psi \in \mathbb{R}^N$  facilitate a remarkable statistical balancing feat, whereby the constrained primal normal eigenaxis components on  $\tau$

$$\tau = \sum_{i=1}^{l_1} \psi_{1i*} \mathbf{x}_{1i*} - \sum_{i=1}^{l_2} \psi_{2i*} \mathbf{x}_{2i*},$$

and  $\tau_0$

$$\tau_0 = \frac{1}{l} \sum_{i=1}^l y_i (1 - \xi_i) - \left( \frac{1}{l} \sum_{i=1}^l \mathbf{x}_{i*} \right)^T \tau,$$

determine a statistical discriminant function

$$D(\mathbf{x}) = \tau^T \mathbf{x} + \tau_0,$$

which delineates a bipartite, symmetric partitioning of a region of large covariance between two overlapping or non-overlapping data distributions in  $\mathbb{R}^d$ . Accordingly, the critical minimum eigenenergies  $\|\tau_1 - \tau_2\|_{\min_c}^2$  of the constrained primal normal eigenlocus components  $\tau_1$  and  $\tau_2$  on  $\tau$  satisfy a linear decision boundary and the bilaterally symmetrical borders which bound it.

It has been demonstrated that the number of constrained primal normal eigenaxis components used to form  $\tau$  is determined by the number of extreme data



points, which are unscaled (unconstrained) primal normal eigenaxis components on  $\tau$ . It has also been demonstrated that any given region of large covariance between two data distributions is a function of extreme training point locations. Previous analyses also indicate that the statistical contents of the Wolfe dual normal eigenaxis components play a significant role in determining a bipartite, symmetric partitioning of a given feature space.

Recall the claim that the total allowed eigenenergy  $\|\tau\|_{\min_c}^2$  of  $\tau$  is the fundamental geometric and statistical property of  $\tau$ , where  $\tau$  possesses a critical minimum eigenenergy

$$\|\tau\|_{\min_c}^2 = \|\tau_1 - \tau_2\|_{\min_c}^2,$$

for which the total allowed eigenenergies of  $\tau_1$  and  $\tau_2$  satisfy a state of statistical equilibrium

$$\left(\|\tau_1\|_{\min_c}^2 - \|\tau_1\| \|\tau_2\| \cos \theta_{\tau_1 \tau_2}\right) + \nabla_{eq} \Leftrightarrow \left(\|\tau_2\|_{\min_c}^2 - \|\tau_1\| \|\tau_2\| \cos \theta_{\tau_1 \tau_2}\right) - \nabla_{eq},$$

in relation to a centrally located statistical fulcrum  $f_s$ , where  $\nabla_{eq}$  is a symmetric equalizer statistic. Recall also that the lengths of the Wolfe dual normal eigenaxis components are selected to satisfy the above state of statistical equilibrium, and thereby balance the total allowed eigenenergies of  $\tau$ .

It will now be shown how the state of equilibrium

$$\sum_{i=1}^{l_1} \psi_{1_{i*}} = \sum_{i=1}^{l_2} \psi_{2_{i*}},$$

that is satisfied by the component lengths of  $\psi$  effectively determines a point of statistical equilibrium

$$\tau = \tau_1 - \tau_2,$$

which exhibits a critical minimum eigenenergy  $\|\tau\|_{\min_c}^2 = \|\tau_1 - \tau_2\|_{\min_c}^2$  that is characterized by joint symmetrical distributions of the eigenenergies of  $\psi$  and  $\tau$ .

Algebraic expressions will now be developed for the total allowed eigenenergies of  $\tau_1$ ,  $\tau_2$ , and  $\tau$ . These expressions will be used to develop statistical equations for the statistical fulcrum  $f_s$  and the symmetric equalizer statistic  $\nabla_{eq}$  in Eq. (34). All of these results will be used to develop the statistical machinery behind the point of statistical equilibrium which is determined by the constrained Lagrangian functional  $L_{\Psi(\tau)}$  of Eq. (15). Figures will be presented that provide geometric and statistical illustrations of the statistical machinery behind the balancing feat.

An algebraic expression will be developed, called the strong dual normal eigenlocus identity, in which the total allowed eigenenergies of a strong dual normal eigenlocus satisfy the law of cosines in surprisingly symmetrical and elegant manners. The strong dual normal eigenlocus identity determines the symmetrical manner in which a strong dual normal eigenlocus  $\tau = \tau_1 - \tau_2$  satisfies a linear decision boundary and the bilaterally symmetrical borders which bound it. All of these results will be used to demonstrate the critical roles that the Wolfe dual normal eigenaxis component lengths  $\left\{\{\psi_{1_{i*}}\}_{i=1}^{l_1}, \{\psi_{2_{i*}}\}_{i=1}^{l_2}\right\}$  and

the equilibrium constraint  $\sum_{i=1}^{l_1} \psi_{1_{i*}} = \sum_{i=1}^{l_2} \psi_{2_{i*}}$  have in determining a bipartite, symmetric partitioning of a given feature space. The analysis begins by returning to the  $\tau_0$  term.

### $\tau_0$ Revisited

Let there be  $l$  extreme training points in a collection of training data. Let there be  $l_1$  eigen-scaled extreme training points  $\psi_{1_{i*}} \mathbf{x}_{1_{i*}}$  that belong to the pattern class  $\mathbf{X}_1$  and  $l_2$  eigen-scaled extreme training points  $\psi_{2_{i*}} \mathbf{x}_{2_{i*}}$  that belong to the pattern class  $\mathbf{X}_2$ . Substituting the expression

$$\tau = \sum_{i=1}^{l_1} \psi_{1_{i*}} \mathbf{x}_{1_{i*}} - \sum_{i=1}^{l_2} \psi_{2_{i*}} \mathbf{x}_{2_{i*}},$$

for  $\tau$  in Eq. (30) into the expression for  $\tau_0$  in Eq. (35) produces the statistic for the  $\tau_0$  term:

$$\begin{aligned} \tau_0 &= \frac{1}{l} \sum_{i=1}^l y_i (1 - \xi_i) \\ &\quad - \frac{1}{l} \sum_{i=1}^l \mathbf{x}_{i*}^T \sum_{j=1}^{l_1} \psi_{1_{j*}} \mathbf{x}_{1_{j*}} \\ &\quad + \frac{1}{l} \sum_{i=1}^l \mathbf{x}_{i*}^T \sum_{j=1}^{l_2} \psi_{2_{j*}} \mathbf{x}_{2_{j*}}. \end{aligned} \tag{77}$$

It will be shown that  $\tau_0$  plays a large role in balancing the total allowed eigenenergies of  $\tau_1$  and  $\tau_2$ . The significance of  $\tau_0$  will be clarified shortly.

The next section examines the symmetrical relationships between the total allowed eigenenergies of  $\tau_1$  and  $\tau_2$ . The demonstration begins with the law of cosines for strong dual normal eigenlocus transforms. The law of cosines determines how a constrained primal normal eigenlocus  $\tau = \tau_1 - \tau_2$  satisfies a linear decision boundary and the bilaterally symmetrical borders which bound it.

## 17.1 The Law of Cosines for Strong Dual Normal Eigenlocus Transforms

The critical minimum eigenenergy  $\|\tau\|_{\min_c}^2$  exhibited by a constrained primal normal eigenlocus  $\tau = \tau_1 - \tau_2$  is regulated by the law of cosines. It will be shown that the law of cosines for strong dual normal eigenlocus transforms requires that the critical minimum eigenenergy  $\|\tau_1\|_{\min_c}^2$  exhibited by the constrained primal normal eigenlocus component  $\tau_1$  coupled with the inner product statistic  $\|\tau_1\| \|\tau_2\| \cos \theta_{\tau_1 \tau_2}$

$$\|\tau_1\|_{\min_c}^2 - \|\tau_1\| \|\tau_2\| \cos \theta_{\tau_1 \tau_2},$$

and the critical minimum eigenenergy  $\|\tau_2\|_{\min_c}^2$  exhibited by the constrained primal normal eigenlocus component  $\tau_2$  coupled with the inner product statistic  $\|\tau_2\| \|\tau_1\| \cos \theta_{\tau_2 \tau_1}$

$$\|\tau_2\|_{\min_c}^2 - \|\tau_2\| \|\tau_1\| \cos \theta_{\tau_2 \tau_1},$$

jointly satisfy the critical minimum eigenenergy  $\|\tau\|_{\min_c}^2$  exhibited by the constrained primal normal eigenlocus  $\tau$

$$\left\{ \|\tau_1\|_{\min_c}^2 - \|\tau_1\| \|\tau_2\| \cos \theta_{\tau_1 \tau_2} \right\} + \left\{ \|\tau_2\|_{\min_c}^2 - \|\tau_2\| \|\tau_1\| \cos \theta_{\tau_2 \tau_1} \right\} = \|\tau\|_{\min_c}^2. \quad (78)$$

Equation (78) indicates that the critical minimum eigenenergy constraints on  $\tau_1$  and  $\tau_2$  include the inner product statistic  $\|\tau_1\| \|\tau_2\| \cos \theta_{\tau_1 \tau_2}$ , which encodes the lengths of  $\tau_1$  and  $\tau_2$  and the angle  $\theta_{\tau_1 \tau_2}$  between them. The sections that follow will demonstrate how all of the constrained primal normal eigenaxis components on a strong dual normal eigenlocus  $\tau = \tau_1 - \tau_2$  satisfy the law of cosines for strong dual normal eigenlocus transforms.

## 17.2 Examining the Total Allowed Eigenenergies of a Strong Dual Normal Eigenlocus

A strong dual normal eigenlocus  $\tau = \tau_1 - \tau_2$  satisfies a linear decision boundary and the bilaterally symmetrical borders which bound it in terms of a critical minimum eigenenergy constraint. The critical minimum eigenenergy exhibited by a constrained primal normal eigenlocus is determined by the KKT constraint of Eq. (21).

Let there be  $l$  eigen-scaled extreme training points on a constrained primal normal eigenlocus  $\tau = \tau_1 - \tau_2$ . The KKT constraint of Eq. (21) requires that the  $l$  constrained primal normal eigenaxis components  $\{\psi_{i_*} \mathbf{x}_{i_*}\}_{i=1}^l$  on  $\tau$  satisfy an algebraic system of  $l$  strong dual normal eigenlocus equations satisfied as strict equalities:

$$\psi_{i_*} \{y_i (\mathbf{x}_{i_*}^T \tau + \tau_0) - 1 + \xi_i\} = 0, \quad i = 1, \dots, l. \quad (79)$$

Equation (79) is now used to examine the critical minimum eigenenergy constraints on  $\tau_1$  and  $\tau_2$ . The analysis begins with the critical minimum eigenenergy constraint on  $\tau_1$ .

### 17.2.1 The Total Allowed Eigenenergy of $\tau_1$

Take any one of the  $l_1$  eigen-scaled extreme training vectors  $\psi_{1_{i_*}} \mathbf{x}_{1_{i_*}}$  that belong to the  $\mathbf{X}_1$  pattern set:  $\{\psi_{1_{i_*}} \mathbf{x}_{1_{i_*}}\}_{i=1}^{l_1}$ . Using Eq. (79) and letting  $y_i = +1$ , it follows that the constrained primal normal eigenaxis component  $\psi_{1_{i_*}} \mathbf{x}_{1_{i_*}}$  on  $\tau_1$  is determined by a strong dual normal eigenlocus equation satisfied as a strict equality:

$$\psi_{1_{i_*}} \left\{ \left( \mathbf{x}_{1_{i_*}}^T \tau + \tau_0 \right) - 1 + \xi_i \right\} = 0,$$

which is part of an algebraic system of  $l_1$  eigenlocus equations.

Now, take all of the  $l_1$  eigen-scaled extreme training vectors that belong to the  $\mathbf{X}_1$  pattern set  $\{\psi_{1_{i_*}} \mathbf{x}_{1_{i_*}}\}_{i=1}^{l_1}$ . Again using Eq. (79) and letting  $y_i = +1$ , it follows that the complete set of  $l_1$  constrained primal normal eigenaxis

components  $\{\psi_{1_{i*}} \mathbf{x}_{1_{i*}}\}_{i=1}^{l_1}$  on  $\tau_1$  is determined by the algebraic system of  $l_1$  strong dual normal eigenlocus equations satisfied as strict equalities:

$$\psi_{1_{i*}} \mathbf{x}_{1_{i*}}^T \tau = \psi_{1_{i*}} (1 - \xi_i - \tau_0), \quad i = 1, \dots, l_1.$$

Using the above expression, it follows that the entire set of  $l_1 \times d$  eigen-transformed extreme vector coordinates of  $\{\psi_{1_{i*}} \mathbf{x}_{1_{i*}}\}_{i=1}^{l_1}$  satisfies the algebraic system of  $l_1$  strong dual normal eigenlocus equations:

$$(1) \quad \psi_{1_{1*}} \mathbf{x}_{1_{1*}}^T \tau = \psi_{1_{1*}} (1 - \xi_1 - \tau_0),$$

$$(2) \quad \psi_{1_{2*}} \mathbf{x}_{1_{2*}}^T \tau = \psi_{1_{2*}} (1 - \xi_2 - \tau_0),$$

$$\vdots$$

$$(l_1) \quad \psi_{1_{l_1*}} \mathbf{x}_{1_{l_1*}}^T \tau = \psi_{1_{l_1*}} (1 - \xi_{l_1} - \tau_0),$$

where each constrained primal normal eigenaxis component  $\psi_{1_{i*}} \mathbf{x}_{1_{i*}}$  on  $\tau_1$  satisfies the statistic:

$$\psi_{1_{i*}} \mathbf{x}_{1_{i*}}^T \tau = \psi_{1_{i*}} (1 - \xi_i - \tau_0).$$

An algebraic expression is now developed for the total allowed eigenenergy of  $\tau_1$ . Denote the total allowed eigenenergy of  $\tau_1$  by  $E_{\tau_1}$  and let  $\tau = \tau_1 - \tau_2$ . Summation over the above algebraic system of  $l_1$  strong dual normal eigenlocus equations produces the following expression for the total allowed eigenenergy  $E_{\tau_1}$  of the constrained primal normal eigenlocus component  $\tau_1$ :

$$\left( \sum_{i=1}^{l_1} \psi_{1_{i*}} \mathbf{x}_{1_{i*}}^T \right) (\tau_1 - \tau_2) \cong \sum_{i=1}^{l_1} \psi_{1_{i*}} (1 - \xi_i - \tau_0),$$

which reduces to

$$\tau_1^T \tau_1 - \tau_1^T \tau_2 \cong \sum_{i=1}^{l_1} \psi_{1_{i*}} (1 - \xi_i - \tau_0),$$

so that the total allowed eigenenergy  $E_{\tau_1}$  of the constrained primal normal eigenlocus component  $\tau_1$  satisfies the equation

$$\|\tau_1\|_{\min_c}^2 - \tau_1^T \tau_2 \cong \sum_{i=1}^{l_1} \psi_{1_{i*}} (1 - \xi_i - \tau_0).$$

It follows that the total allowed eigenenergy  $E_{\tau_1}$  of the constrained primal normal eigenlocus component  $\tau_1$  is determined by the expression

$$\|\tau_1\|_{\min_c}^2 - \|\tau_1\| \|\tau_2\| \cos \theta_{\tau_1 \tau_2} \cong \sum_{i=1}^{l_1} \psi_{1_{i*}} (1 - \xi_i - \tau_0), \quad (80)$$

where the total allowed eigenenergy  $E_{\tau_1}$  of  $\tau_1$

$$E_{\tau_1} \triangleq \|\tau_1\|_{\min_c}^2 - \|\tau_1\| \|\tau_2\| \cos \theta_{\tau_1 \tau_2},$$

is determined by a statistical equation

$$E_{\tau_1} = \sum_{i=1}^{l_1} \psi_{1_{i*}} (1 - \xi_i - \tau_0),$$

in terms of integrated magnitudes  $\sum_{i=1}^{l_1} \psi_{1_{i*}}$  of Wolfe dual normal eigenaxis components which are correlated with extreme vectors that belong to the  $\mathbf{X}_1$  pattern set, and the  $\tau_0$  statistic. The critical minimum eigenenergy constraint on  $\tau_2$  is examined next.

### 17.2.2 The Total Allowed Eigenenergy of $\tau_2$

Take any one of the  $l_2$  eigen-scaled extreme training vectors  $\psi_{2_{i*}} \mathbf{x}_{2_{i*}}$  that belong to the  $\mathbf{X}_2$  pattern set:  $\{\psi_{2_{i*}} \mathbf{x}_{2_{i*}}\}_{i=1}^{l_2}$ . Using Eq. (79) and letting  $y_i = -1$ , it follows that a constrained primal normal eigenaxis component  $\psi_{2_{i*}} \mathbf{x}_{2_{i*}}$  on  $\tau_2$  is determined by a strong dual normal eigenlocus equation satisfied as a strict equality:

$$\psi_{2_{i*}} \left\{ \left( -\mathbf{x}_{2_{i*}}^T \tau - \tau_0 \right) - 1 + \xi_i \right\} = 0,$$

which is part of an algebraic system of  $l_2$  eigenlocus equations.

Now, take all of the  $l_2$  eigen-scaled extreme training vectors that belong to the  $\mathbf{X}_2$  pattern set:  $\{\psi_{2_{i*}} \mathbf{x}_{2_{i*}}\}_{i=1}^{l_2}$ . Again using Eq. (79) and letting  $y_i = -1$ , it follows that the complete set of  $l_2$  constrained primal normal eigenaxis components  $\{\psi_{2_{i*}} \mathbf{x}_{2_{i*}}\}_{i=1}^{l_2}$  on  $\tau_2$  is determined by the algebraic system of  $l_2$  strong dual normal eigenlocus equations satisfied as strict equalities:

$$-\psi_{2_{i*}} \mathbf{x}_{2_{i*}}^T \tau = \psi_{2_{i*}} (1 - \xi_i + \tau_0), \quad i = 1, \dots, l_2.$$

Using the above expression, it follows that the entire set of  $l_2 \times d$  eigen-transformed extreme vector coordinates of  $\{\psi_{2_{i*}} \mathbf{x}_{2_{i*}}\}_{i=1}^{l_2}$  satisfies the algebraic system of  $l_2$  strong dual normal eigenlocus equations:

$$\begin{aligned} (1) \quad & -\psi_{2_{1*}} \mathbf{x}_{2_{1*}}^T \tau = \psi_{2_{1*}} (1 - \xi_1 + \tau_0), \\ (2) \quad & -\psi_{2_{2*}} \mathbf{x}_{2_{2*}}^T \tau = \psi_{2_{2*}} (1 - \xi_2 + \tau_0), \\ & \vdots \\ (l_2) \quad & -\psi_{2_{l_2*}} \mathbf{x}_{2_{l_2*}}^T \tau = \psi_{2_{l_2*}} (1 - \xi_{l_2} + \tau_0), \end{aligned}$$

where each constrained primal normal eigenaxis component  $\psi_{2_{i*}} \mathbf{x}_{2_{i*}}$  on  $\tau_2$  satisfies the statistic:

$$-\psi_{2_{i*}} \mathbf{x}_{2_{i*}}^T \tau = \psi_{2_{i*}} (1 - \xi_i + \tau_0).$$

An algebraic expression is now developed for the total allowed eigenenergy of  $\tau_2$ . Denote the total allowed eigenenergy of  $\tau_2$  by  $E_{\tau_2}$  and let  $\tau = \tau_1 - \tau_2$ . Summation over the above algebraic system of  $l_2$  strong dual normal eigenlocus

equations produces the following expression for the total allowed eigenenergy  $E_{\tau_2}$  of the constrained primal normal eigenlocus component  $\tau_2$ :

$$-\left(\sum_{i=1}^{l_2} \psi_{2i_*} \mathbf{x}_{2i_*}^T\right) (\tau_1 - \tau_2) \cong \sum_{i=1}^{l_2} \psi_{2i_*} (1 - \xi_i + \tau_0),$$

which reduces to

$$\tau_2^T \tau_2 - \tau_2^T \tau_1 \cong \sum_{i=1}^{l_2} \psi_{2i_*} (1 - \xi_i + \tau_0),$$

so that the total allowed eigenenergy  $E_{\tau_2}$  of the constrained primal normal eigenlocus component  $\tau_2$  satisfies the equation

$$\|\tau_2\|_{\min_c}^2 - \tau_2^T \tau_1 \cong \sum_{i=1}^{l_2} \psi_{2i_*} (1 - \xi_i + \tau_0).$$

It follows that the total allowed eigenenergy  $E_{\tau_2}$  of the constrained primal normal eigenlocus component  $\tau_2$  is determined by the expression

$$\|\tau_2\|_{\min_c}^2 - \|\tau_2\| \|\tau_1\| \cos \theta_{\tau_2 \tau_1} \cong \sum_{i=1}^{l_2} \psi_{2i_*} (1 - \xi_i + \tau_0), \quad (81)$$

where the total allowed eigenenergy  $E_{\tau_2}$  of  $\tau_2$

$$E_{\tau_2} \triangleq \|\tau_2\|_{\min_c}^2 - \|\tau_2\| \|\tau_1\| \cos \theta_{\tau_2 \tau_1},$$

is determined by a statistical equation

$$E_{\tau_2} = \sum_{i=1}^{l_2} \psi_{2i_*} (1 - \xi_i + \tau_0),$$

in terms of integrated magnitudes  $\sum_{i=1}^{l_2} \psi_{2i_*}$  of Wolfe dual normal eigenaxis components which are correlated with extreme vectors that belong to the  $\mathbf{X}_2$  pattern set, and the  $\tau_0$  statistic. An algebraic expression is now obtained for the total allowed eigenenergy of a constrained primal normal eigenlocus  $\tau$ . Denote the total allowed eigenenergy  $\|\tau\|_{\min_c}^2$  of  $\tau$  by  $E_\tau$ .

### 17.2.3 The Total Allowed Eigenenergy of $\tau$

Summation over the complete algebraic system of strong dual normal eigenlocus equations satisfied by  $\tau_1$

$$\left(\sum_{i=1}^{l_1} \psi_{1i_*} \mathbf{x}_{1i_*}^T\right) \tau \cong \sum_{i=1}^{l_1} \psi_{1i_*} (1 - \xi_i - \tau_0),$$

and by  $\tau_2$

$$\left(-\sum_{i=1}^{l_2} \psi_{2i_*} \mathbf{x}_{2i_*}^T\right) \tau \cong \sum_{i=1}^{l_2} \psi_{2i_*} (1 - \xi_i + \tau_0),$$

produces the following expression for the total allowed eigenenergy  $E_\tau$  of  $\tau$

$$\begin{aligned} & \left( \sum_{i=1}^{l_1} \psi_{1_{i*}} \mathbf{x}_{1_{i*}}^T - \sum_{i=1}^{l_2} \psi_{2_{i*}} \mathbf{x}_{2_{i*}}^T \right) \tau \\ & \cong \sum_{i=1}^{l_1} \psi_{1_{i*}} (1 - \xi_i - \tau_0) + \sum_{i=1}^{l_2} \psi_{2_{i*}} (1 - \xi_i + \tau_0), \end{aligned}$$

which reduces to

$$\begin{aligned} (\tau_1 - \tau_2)^T \tau & \cong \sum_{i=1}^{l_1} \psi_{1_{i*}} (1 - \xi_i - \tau_0) \\ & + \sum_{i=1}^{l_2} \psi_{2_{i*}} (1 - \xi_i + \tau_0), \\ & \cong \sum_{i=1}^l \psi_{i*} (1 - \xi_i), \end{aligned}$$

so that the total allowed eigenenergy  $E_\tau$  of  $\tau$

$$(\tau_1 - \tau_2)^T \tau = \|\tau\|_{\min_c}^2,$$

is determined by a statistical equation

$$E_\tau = \sum_{i=1}^l \psi_{i*} (1 - \xi_i),$$

solely in terms of the integrated lengths of the Wolfe dual normal eigenaxis components on  $\psi$ . It follows that the total allowed eigenenergy  $E_\tau$  of a constrained primal normal eigenlocus  $\tau$  is determined by the integrated magnitudes  $\psi_{i*}$  of the Wolfe dual normal eigenaxis components  $\psi_{i*} \vec{\mathbf{e}}_{i*}$  on  $\psi$

$$\begin{aligned} \|\tau\|_{\min_c}^2 & \cong \sum_{i=1}^l \psi_{i*} (1 - \xi_i) \\ & \cong \sum_{i=1}^l \psi_{i*} - \sum_{i=1}^l \psi_{i*} \xi_i, \end{aligned} \tag{82}$$

where the regularization parameters  $\xi_i = \xi \ll 1$  are seen to determine negligible constraints on  $E_\tau$ .

The next part of the analysis will identify the manner in which the total allowed eigenenergies of  $\tau_1$  and  $\tau_2$  are symmetrically balanced with each other.

### 17.3 Balancing the Total Allowed Eigenenergies of $\tau_1$ and $\tau_2$

Returning to Eq. (34), recall that the critical minimum eigenenergies of  $\tau_1$  and  $\tau_2$  satisfy a state of statistical equilibrium

$$\left( \|\tau_1\|_{\min_c}^2 - \|\tau_1\| \|\tau_2\| \cos \theta_{\tau_1 \tau_2} \right) + \nabla_{eq} \Leftrightarrow \left( \|\tau_2\|_{\min_c}^2 - \|\tau_1\| \|\tau_2\| \cos \theta_{\tau_1 \tau_2} \right) - \nabla_{eq},$$

in relation to a centrally located statistical fulcrum  $f_s$ , where  $\nabla_{eq}$  is a symmetric equalizer statistic. Statistical expressions are now obtained for the symmetric equalizer statistic  $\nabla_{eq}$  and the statistical fulcrum  $f_s$ .

Using Eq. (80) and the KKT constraint of Eq. (38)

$$\begin{aligned}\|\tau_1\|_{\min_c}^2 - \|\tau_1\| \|\tau_2\| \cos \theta_{\tau_1 \tau_2} &\cong \sum_{i=1}^{l_1} \psi_{1_{i*}} (1 - \xi_i - \tau_0), \\ &\cong \frac{1}{2} \sum_{i=1}^l \psi_{i*} (1 - \xi_i - \tau_0),\end{aligned}$$

it follows that the total allowed eigenenergy  $E_{\tau_1}$  of  $\tau_1$

$$E_{\tau_1} \triangleq \|\tau_1\|_{\min_c}^2 - \|\tau_1\| \|\tau_2\| \cos \theta_{\tau_1 \tau_2},$$

is constrained to satisfy the statistical equation:

$$E_{\tau_1} = \frac{1}{2} \sum_{i=1}^l \psi_{i*} (1 - \xi_i) - \frac{\tau_0}{2} \sum_{i=1}^l \psi_{i*}. \quad (83)$$

Using Eq. (81) and the KKT constraint of Eq. (38)

$$\begin{aligned}\|\tau_2\|_{\min_c}^2 - \|\tau_2\| \|\tau_1\| \cos \theta_{\tau_2 \tau_1} &\cong \sum_{i=1}^{l_2} \psi_{2_{i*}} (1 - \xi_i + \tau_0), \\ &\cong \frac{1}{2} \sum_{i=1}^l \psi_{i*} (1 - \xi_i + \tau_0),\end{aligned}$$

it follows that the total allowed eigenenergy  $E_{\tau_2}$  of  $\tau_2$

$$E_{\tau_2} \triangleq \|\tau_2\|_{\min_c}^2 - \|\tau_2\| \|\tau_1\| \cos \theta_{\tau_2 \tau_1},$$

is constrained to satisfy the statistical equation:

$$E_{\tau_2} = \frac{1}{2} \sum_{i=1}^l \psi_{i*} (1 - \xi_i) + \frac{\tau_0}{2} \sum_{i=1}^l \psi_{i*}. \quad (84)$$

Using Eqs (82), (83), and (84), it follows that the total allowed eigenenergies of the constrained primal normal eigenaxis components on  $\tau_1$

$$\begin{aligned}E_{\tau_1} &= \frac{1}{2} \sum_{i=1}^l \psi_{i*} (1 - \xi_i) - \frac{\tau_0}{2} \sum_{i=1}^l \psi_{i*}, \\ &\cong \frac{1}{2} \|\tau\|_{\min_c}^2 (1 - \xi_i) - \frac{\tau_0}{2} \|\tau\|_{\min_c}^2,\end{aligned}$$

and the total allowed eigenenergies of the constrained primal normal eigenaxis components on  $\tau_2$

$$\begin{aligned}E_{\tau_2} &= \frac{1}{2} \sum_{i=1}^l \psi_{i*} (1 - \xi_i) + \frac{\tau_0}{2} \sum_{i=1}^l \psi_{i*}, \\ &\cong \frac{1}{2} \|\tau\|_{\min_c}^2 (1 - \xi_i) + \frac{\tau_0}{2} \|\tau\|_{\min_c}^2,\end{aligned}$$

are symmetrically balanced with each other by means of the symmetric equalizer statistic  $\nabla_{eq}$ :

$$\nabla_{eq} = \frac{\tau_0}{2} \|\tau\|_{\min_c}^2, \quad (85)$$



in relation to the centrally located statistical fulcrum  $f_s$ :

$$\begin{aligned} f_s &= \frac{1}{2} \sum_{i=1}^l \psi_{i*} (1 - \xi_i), \\ &\approx \frac{1}{2} \|\tau\|_{\min_c}^2, \end{aligned} \quad (86)$$

which is half the total allowed eigenenergy  $\frac{1}{2}E_\tau$  of a strong dual normal eigenlocus  $\tau$ .

Equations (85) and (86) are now used to define the state of statistical equilibrium of a strong dual normal eigenlocus.

#### 17.4 The State of Statistical Equilibrium of a Strong Dual Normal Eigenlocus

Let  $E_{\tau_1} \triangleq \|\tau_1\|_{\min_c}^2 - \|\tau_1\| \|\tau_2\| \cos \theta_{\tau_1 \tau_2}$  and  $E_{\tau_2} \triangleq \|\tau_2\|_{\min_c}^2 - \|\tau_2\| \|\tau_1\| \cos \theta_{\tau_2 \tau_1}$ . Using Eqs (34), (85), and (86), it follows that the total allowed eigenenergies of  $\tau_1$  and  $\tau_2$  satisfy the state of statistical equilibrium

$$E_{\tau_1} + \frac{\tau_0}{2} \|\tau\|_{\min_c}^2 \Leftrightarrow E_{\tau_2} - \frac{\tau_0}{2} \|\tau\|_{\min_c}^2, \quad (87)$$

in relation to the centrally located statistical fulcrum  $f_s$

$$\begin{aligned} f_s &= \frac{1}{2} \sum_{i=1}^l \psi_{i*}, \\ &\cong \frac{1}{2} \|\tau\|_{\min_c}^2, \end{aligned}$$

which is located at half the critical eigenenergy  $\frac{1}{2}E_\tau$  of a strong dual normal eigenlocus  $\tau$ . Clearly, then, the  $\tau_0$  term plays a significant role in balancing the total allowed eigenenergies of  $\tau_1$  and  $\tau_2$ . The geometric and statistical meaning of Eq. (87) is considered next.

#### 17.5 The Statistical Machinery Behind the Balancing Feat

It has been shown that correlated normal eigenaxis components on  $\psi$  and  $\tau$  exhibit directional symmetry. It has also been shown that joint distributions of normal eigenaxis components on  $\psi$  and  $\tau$  are symmetrically distributed over the Wolfe dual normal eigenaxis components  $\psi_{1i*} \vec{e}_{1i*}$  and  $\psi_{2i*} \vec{e}_{2i*}$  on  $\psi$  and the constrained primal normal eigenaxis components  $\psi_{1i*} \mathbf{x}_{1i*}$  and  $\psi_{2i*} \mathbf{x}_{2i*}$  on  $\tau_1$  and  $\tau_2$ . In addition, the previous analysis has demonstrated how strong duality relationships between the component lengths of  $\psi$  and  $\tau$  ensure that products of joint component lengths, i.e., the total allowed eigenenergies  $\|\psi_{1i*} \mathbf{x}_{1i*}\|_{\min_c}^2$  and  $\|\psi_{2i*} \mathbf{x}_{2i*}\|_{\min_c}^2$  of the constrained primal normal eigenaxis components  $\psi_{1i*} \mathbf{x}_{1i*}$  and  $\psi_{2i*} \mathbf{x}_{2i*}$ , are symmetrically balanced with each other.

A geometric and statistical explanation of Eq. (87) is now obtained. The explanation uses the general machinery of a fulcrum and a lever, where a lever

is any rigid object capable of turning about some fixed point called a fulcrum. If a fulcrum is placed directly under a lever's center of gravity, the lever will remain balanced. If a lever is of uniform dimensions and density, then the center of gravity is at the geometric center of the lever. Consider for example, the playground device known as a seesaw or teeter-totter. The center of gravity is at the geometric center of a teeter-totter, which is where the fulcrum of a seesaw is located Asimov [1966].

## 17.6 Statistical Machinery of a Statistical Fulcrum and a Statistical Lever

Consider an implicit horizontal axis capable of rotating about some fixed point, where the horizontal axis is a statistical lever, and the fixed point is a centrally located statistical fulcrum. Let the joint eigenenergies of  $\psi$  and  $\tau$  be symmetrically distributed over the statistical lever, and let the statistical fulcrum  $f_s$  be located directly under the statistical lever's center of eigenenergy, which is defined to be half the critical minimum eigenenergy  $\frac{1}{2} \|\tau\|_{\min_c}^2$  of a strong dual normal eigenlocus  $\tau$ . Using Eq. (87), let the statistical lever be subjected to equal and opposite eigenenergies associated with  $\tau$ ,  $\tau_1$ , and  $\tau_2$ , in terms of  $E_{\tau_1} + \frac{\tau_0}{2} \|\tau\|_{\min_c}^2$  and  $E_{\tau_2} - \frac{\tau_0}{2} \|\tau\|_{\min_c}^2$ , where the symmetric equalizer statistic  $\nabla_{eq} = \frac{\tau_0}{2} \|\tau\|_{\min_c}^2$  ensures that the total allowed eigenenergy  $E_{\tau_1}$  of  $\tau_1$  and the total allowed eigenenergy  $E_{\tau_2}$  of  $\tau_2$  are symmetrically balanced with each other. Given that equal and opposite eigenenergies are applied to the statistical lever, it follows that the statistical lever achieves a state of statistical equilibrium.

Figure 29 illustrates the elegant statistical machinery which ensures that the total allowed eigenenergies of  $\tau_1$  and  $\tau_2$  are symmetrically balanced with each other. The statistical lever is depicted by a gray bar and the statistical fulcrum is depicted by a purple triangle. The statistical lever, which pivots about the statistical fulcrum, is subjected to equal and opposite eigenenergies associated with  $\tau$ ,  $\tau_1$ , and  $\tau_2$ , and therefore achieves a state of statistical equilibrium.

## Equal and Opposite Eigenenergies

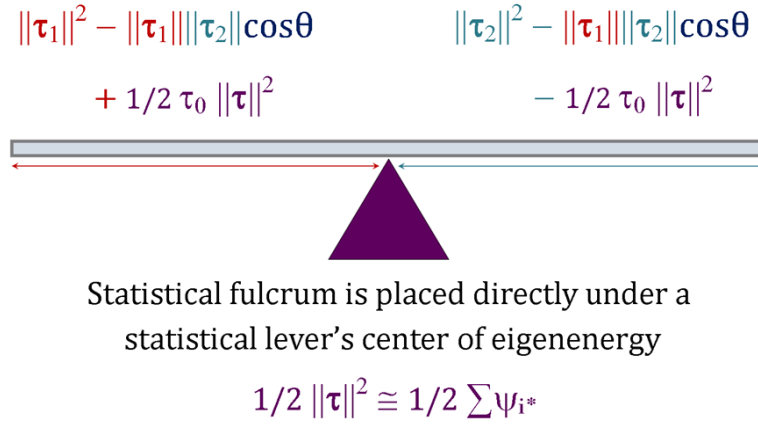


Figure 29: Illustration of the statistical machinery that is used to symmetrically balance the eigenenergies of  $\tau_1$  and  $\tau_2$ . The statistical lever, which is subjected to equal and opposite eigenenergies of  $\tau$ ,  $\tau_1$ , and  $\tau_2$ , achieves a state of statistical equilibrium.

### 17.7 Characteristics of the State of Statistical Equilibrium

The state of statistical equilibrium that is achieved by a strong dual normal eigenlocus results from the total allowed eigenenergy  $\|\tau\|_{\min_c}^2 \cong \|\tau_1 - \tau_2\|_{\min_c}^2$  of  $\tau$  and the total allowed eigenenergy  $\|\psi\|_{\min_c}^2$  of  $\psi$  being determined by joint symmetrical distributions of the eigenenergies of  $\psi$  and  $\tau$ , whereby the eigenenergies of the eigen-scaled extreme training points on  $\tau_1$  and  $\tau_2$  are distributed in a manner which symmetrically balances the eigenenergies of  $\tau_1$  and  $\tau_2$ . Thereby, the total allowed eigenenergy  $\|\tau\|_{\min_c}^2 \cong \|\tau_1 - \tau_2\|_{\min_c}^2$  of  $\tau$  satisfies a linear decision boundary and the bilaterally symmetrical borders which bound it.

The state of statistical equilibrium is characterized by joint symmetrical distributions of the normal eigenaxis components on  $\psi$  and  $\tau$ , over all of the eigen-scaled extreme training points on  $\tau_1$  and  $\tau_2$ , whereby symmetrical distributions of the critical minimum eigenenergies of  $\psi$  and  $\tau$  jointly satisfy the law of cosines for strong dual normal eigenlocus transforms in Eq. (78), in symmetrical and elegant manners.

Indeed, *all of the eigen-scaled extreme points on  $\tau_1$  and  $\tau_2$  possess such eigen-balanced locations*, that the inner product statistic  $\|\tau_1\| \|\tau_2\| \cos \theta_{\tau_1 \tau_2}$  coupled with the critical minimum eigenenergy  $\|\tau_1\|_{\min_c}^2$  of the constrained primal

normal eigenlocus component  $\tau_1$

$$\|\tau_1\|_{\min_c}^2 - \|\tau_1\| \|\tau_2\| \cos \theta_{\tau_1 \tau_2},$$

and the inner product statistic  $\|\tau_2\| \|\tau_1\| \cos \theta_{\tau_2 \tau_1}$  coupled with the critical minimum eigenenergy  $\|\tau_2\|_{\min_c}^2$  of the constrained primal normal eigenlocus component  $\tau_2$

$$\|\tau_2\|_{\min_c}^2 - \|\tau_2\| \|\tau_1\| \cos \theta_{\tau_2 \tau_1},$$

are effectively balanced with the critical minimum eigenenergy  $\|\tau\|_{\min_c}^2$  of the constrained primal normal eigenlocus  $\tau$ :

$$\left\{ \|\tau_1\|_{\min_c}^2 - \|\tau_1\| \|\tau_2\| \cos \theta_{\tau_1 \tau_2} \right\} + \left\{ \|\tau_2\|_{\min_c}^2 - \|\tau_2\| \|\tau_1\| \cos \theta_{\tau_1 \tau_2} \right\} \cong \|\tau\|_{\min_c}^2.$$

Correspondingly, *all of the Wolfe dual normal eigenaxis components on  $\psi$  possess such eigen-balanced magnitudes*, that component lengths of  $\psi$  which regulate the critical minimum eigenenergy of  $\tau_1$

$$E_{\tau_1} = \sum_{i=1}^{l_1} \psi_{1_{i*}} - \tau_0 \sum_{i=1}^{l_1} \psi_{1_{i*}},$$

and component lengths of  $\psi$  which regulate the critical minimum eigenenergy of  $\tau_2$

$$E_{\tau_2} = \sum_{i=1}^{l_2} \psi_{2_{i*}} + \tau_0 \sum_{i=1}^{l_2} \psi_{2_{i*}},$$

are effectively balanced with component lengths of  $\psi$  which regulate the critical minimum eigenenergy of  $\tau$ :

$$\begin{aligned} E_{\tau} &= \sum_{i=1}^{l_1} \psi_{1_{i*}} (1 - \tau_0) + \sum_{i=1}^{l_2} \psi_{2_{i*}} (1 + \tau_0), \\ &\cong \sum_{i=1}^{l_1} \psi_{1_{i*}} + \sum_{i=1}^{l_2} \psi_{2_{i*}}, \\ &\cong \frac{1}{2} \sum_{i=1}^l \psi_{i*} + \frac{1}{2} \sum_{i=1}^l \psi_{i*}, \\ &\cong \sum_{i=1}^l \psi_{i*}. \end{aligned}$$

Figure 30 illustrates how the statistical equilibrium point of a strong dual normal eigenlocus  $\tau$  is characterized by joint symmetrical distributions of the critical minimum eigenenergies of  $\psi$  and  $\tau$  which jointly satisfy the law of cosines for strong dual normal eigenlocus transforms. Thereby, the eigenenergies of the eigen-scaled extreme training points on  $\tau_1$  and  $\tau_2$  are distributed in a symmetric manner which symmetrically balances the eigenenergies of  $\tau_1$  and  $\tau_2$ .

## Statistical Equilibrium Point of $\tau$

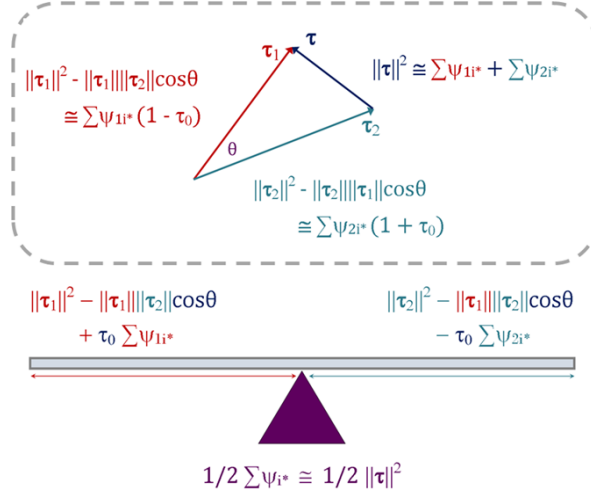


Figure 30: Illustration of the geometric and topological relationships which determine the critical minimum eigenenergies and component lengths of a Wolfe dual and a constrained primal normal eigenlocus. Figure 30 shows how the statistical equilibrium point (the dual statistical eigenlocus) of  $\tau$  is determined by symmetrically balanced eigenenergies of the constrained pair of primal normal eigenlocus components  $\tau_1 - \tau_2$  on  $\tau$ . The critical minimum eigenenergies of  $\psi$  and  $\tau$  jointly satisfy the law of cosines for strong dual normal eigenlocus transforms.

The law of cosines which is satisfied by strong dual normal eigenlocus transforms will be formally referred to as *the strong dual normal eigenlocus identity*. The components on a Wolfe dual normal eigenlocus  $\psi$  and a constrained primal normal eigenlocus  $\tau = \tau_1 - \tau_2$  satisfy the strong dual normal eigenlocus identity that is outlined next. Without loss of generality, regularization parameters  $\xi_i$  are not included in the identity.

### 17.8 The Strong Dual Normal Eigenlocus Identity

Figure 30 shows that the total allowed eigenenergies  $\|\tau\|_{\min_c}^2 \cong \|\tau_1 - \tau_2\|_{\min_c}^2$  of a strong dual normal eigenlocus  $\tau = \tau_1 - \tau_2$

$$\begin{aligned} \|\tau\|_{\min_c}^2 &\cong \|\tau_1 - \tau_2\|_{\min_c}^2, \\ &\cong \left\{ \|\tau_1\|_{\min_c}^2 - \|\tau_1\| \|\tau_2\| \cos \theta_{\tau_1 \tau_2} \right\} + \left\{ \|\tau_2\|_{\min_c}^2 - \|\tau_2\| \|\tau_1\| \cos \theta_{\tau_1 \tau_2} \right\}, \end{aligned}$$

and the corresponding total allowed eigenenergy of  $\|\psi\|_{\min_c}^2$  of a Wolfe dual normal eigenlocus  $\psi$ , *jointly satisfy the law of cosines in a surprisingly elegant*

and symmetric manner

$$\begin{aligned}
\|\tau\|_{\min_c}^2 &\cong \left\{ \|\tau_1\|_{\min_c}^2 - \|\tau_1\| \|\tau_2\| \cos \theta_{\tau_1 \tau_2} \right\} + \left\{ \|\tau_2\|_{\min_c}^2 - \|\tau_2\| \|\tau_1\| \cos \theta_{\tau_1 \tau_2} \right\}, \\
&\cong \sum_{i=1}^{l_1} \psi_{1_{i_*}} (1 - \tau_0) + \sum_{i=1}^{l_2} \psi_{2_{i_*}} (1 + \tau_0), \\
&\cong \sum_{i=1}^{l_1} \psi_{1_{i_*}} + \sum_{i=1}^{l_2} \psi_{2_{i_*}}, \\
&\cong \frac{1}{2} \|\tau\|_{\min_c}^2 + \frac{1}{2} \|\tau\|_{\min_c}^2, \\
&\cong \frac{1}{2} \sum_{i=1}^l \psi_{i_*} + \frac{1}{2} \sum_{i=1}^l \psi_{i_*}, \\
&\cong \|\tau_1 - \tau_2\|_{\min_c}^2, \\
&\cong \sum_{i=1}^l \psi_{i_*}.
\end{aligned} \tag{88}$$

Equation (88) is given the name of the strong dual normal eigenlocus identity. Figure 30 and Eq. (88) illustrate the algebraic and geometric nature of the symmetrical relationships between the total allowed eigenenergies of the components of  $\tau$  and the components of  $\psi$ .

Equation (88) shows how the total allowed eigenenergy  $E_\tau$  of a strong dual normal eigenlocus  $\tau$  is regulated by the integrated component lengths  $\sum_{i=1}^l \psi_{i_*}$  of a Wolfe dual normal eigenlocus  $\psi$

$$\begin{aligned}
\|\tau\|_{\min_c}^2 &\cong \sum_{i=1}^l \psi_{i_*}, \\
&\cong 2 \sum_{i=1}^{l_1} \psi_{1_{i_*}}, \\
&\cong 2 \sum_{i=1}^{l_2} \psi_{2_{i_*}},
\end{aligned}$$

where  $\sum_{i=1}^{l_1} \psi_{1_{i_*}} \equiv \sum_{i=1}^{l_2} \psi_{2_{i_*}}$ , in terms of symmetrically weighted, symmetrical sets of integrated component lengths

$$\begin{aligned}
\|\tau\|_{\min_c}^2 &\cong \sum_{i=1}^{l_1} \psi_{1_{i_*}} - \tau_0 \sum_{i=1}^{l_1} \psi_{1_{i_*}} \\
&\quad + \sum_{i=1}^{l_2} \psi_{2_{i_*}} + \tau_0 \sum_{i=1}^{l_2} \psi_{2_{i_*}},
\end{aligned}$$

where the symmetrically weighted, integrated component lengths  $\frac{1}{2} (1 - \tau_0) \sum_{i=1}^l \psi_{i_*}$  of  $\psi$  which regulate the total allowed eigenenergy  $E_{\tau_1}$  of  $\tau_1$

$$\begin{aligned}
E_{\tau_1} &= \sum_{i=1}^{l_1} \psi_{1_{i_*}} (1 - \tau_0), \\
&\cong \frac{1}{2} (1 - \tau_0) \sum_{i=1}^l \psi_{i_*}, \\
&\cong \frac{1}{2} (1 - \tau_0) \|\tau\|_{\min_c}^2,
\end{aligned}$$

and the symmetrically weighted, integrated component lengths  $\frac{1}{2} (1 + \tau_0) \sum_{i=1}^l \psi_{i_*}$  of  $\psi$  which regulate the total allowed eigenenergy  $E_{\tau_2}$  of  $\tau_2$

$$\begin{aligned} E_{\tau_2} &= \sum_{i=1}^{l_2} \psi_{2i_*} (1 + \tau_0), \\ &\cong \frac{1}{2} (1 + \tau_0) \sum_{i=1}^l \psi_{i_*}, \\ &\cong \frac{1}{2} (1 + \tau_0) \|\tau\|_{\min_c}^2, \end{aligned}$$

determine symmetrically balanced, critical minimum eigenenergies that sum to the total allowed eigenenergy  $E_\tau$  of  $\tau$

$$\begin{aligned} E_{\tau_1} + E_{\tau_2} &= \sum_{i=1}^{l_1} \psi_{1i_*} (1 - \tau_0) + \sum_{i=1}^{l_2} \psi_{2i_*} (1 + \tau_0), \\ &\cong \frac{1}{2} (1 - \tau_0) \sum_{i=1}^l \psi_{i_*} + \frac{1}{2} (1 + \tau_0) \sum_{i=1}^l \psi_{i_*}, \\ &\cong \frac{1}{2} (1 - \tau_0) \|\tau\|_{\min_c}^2 + \frac{1}{2} (1 + \tau_0) \|\tau\|_{\min_c}^2, \\ &\cong \|\tau\|_{\min_c}^2, \\ &= E_\tau. \end{aligned}$$

It has previously been shown that joint distributions of the normal eigenaxis components on  $\psi$  and  $\tau$  are symmetrically distributed over the Wolfe dual normal eigenaxis  $\psi_{1i_*} \vec{e}_{1i_*}$  and  $\psi_{2i_*} \vec{e}_{2i_*}$  on  $\psi$  and the constrained primal normal eigenaxis components  $\psi_{1i_*} \mathbf{x}_{1i_*}$  and  $\psi_{2i_*} \mathbf{x}_{2i_*}$  on  $\tau_1$  and  $\tau_2$ . Figure 30 and Eq. (88) illustrate how the total allowed eigenenergies of the constrained primal normal eigenaxis components  $\psi_{1i_*} \mathbf{x}_{1i_*}$  and  $\psi_{2i_*} \mathbf{x}_{2i_*}$  on  $\tau = \tau_1 - \tau_2$  are regulated by symmetrical relationships between the total allowed eigenenergies of  $\psi$  and  $\tau$ . Indeed, the total allowed eigenenergies of the constrained primal normal eigenaxis components on  $\tau_1$

$$\left\| \sum_{i=1}^{l_1} \psi_{1i_*} \mathbf{x}_{1i_*} \right\|_{\min_c}^2 - \left\| \sum_{i=1}^{l_1} \psi_{1i_*} \mathbf{x}_{1i_*} \right\| \left\| \sum_{i=1}^{l_2} \psi_{2i_*} \mathbf{x}_{2i_*} \right\| \cos \theta_{\tau_1 \tau_2},$$

are regulated by the statistical equations

$$\begin{aligned} E_{\tau_1} &= \sum_{i=1}^{l_1} \psi_{1i_*} (1 - \tau_0), \\ &\cong \frac{1}{2} (1 - \tau_0) \sum_{i=1}^l \psi_{i_*}, \\ &\cong \frac{1}{2} (1 - \tau_0) \|\tau\|_{\min_c}^2, \end{aligned}$$

and the total allowed eigenenergies of the constrained primal normal eigenaxis components on  $\tau_2$

$$\left\| \sum_{i=1}^{l_2} \psi_{2i_*} \mathbf{x}_{2i_*} \right\|_{\min_c}^2 - \left\| \sum_{i=1}^{l_1} \psi_{1i_*} \mathbf{x}_{1i_*} \right\| \left\| \sum_{i=1}^{l_2} \psi_{2i_*} \mathbf{x}_{2i_*} \right\| \cos \theta_{\tau_1 \tau_2},$$

are regulated by the statistical equations

$$\begin{aligned} E_{\tau_2} &= \sum_{i=1}^{l_2} \psi_{2_{i_*}} (1 + \tau_0), \\ &\cong \frac{1}{2} (1 + \tau_0) \sum_{i=1}^l \psi_{i_*}, \\ &\cong \frac{1}{2} (1 + \tau_0) \|\tau\|_{\min_c}^2, \end{aligned}$$

whereby the total allowed eigenenergies of the constrained primal normal eigenaxis components on  $\tau_1$  and  $\tau_2$

$$\left\| \sum_{i=1}^{l_1} \psi_{1_{i_*}} \mathbf{x}_{1_{i_*}} - \sum_{i=1}^{l_2} \psi_{2_{i_*}} \mathbf{x}_{2_{i_*}} \right\|_{\min_c}^2,$$

are determined by joint symmetrical distributions of the eigenenergies of  $\psi$  and  $\tau$ .

## Configurations of Regularized Geometric Architectures

The learning machine architecture which has been examined in this paper exhibits a surprising amount of bilateral symmetry for arbitrary data distributions. Given the previous analyses, it is concluded that robust and symmetrical linear partitions of arbitrary feature spaces are achieved by means of symmetrically balanced normal eigenaxis components in dual, correlated Hilbert spaces.

Returning to Figs. 12 and 20, it is concluded that the regularized, data-driven geometric architectures determined by strong dual normal eigenlocus transforms are configured by enforcing joint symmetrical distributions of the eigenenergies of  $\psi$  and  $\tau$  over the eigen-scaled extreme training vectors on  $\tau_1$  and  $\tau_2$ , whereby the eigenenergies of the strong dual normal eigenlocus components  $\tau_1$  and  $\tau_2$  on  $\tau$  are symmetrically balanced with each other.

Figure 31 illustrates how the joint eigenenergies of  $\psi$  and  $\tau$  are symmetrically distributed over the constrained primal normal eigenaxis components on the constrained primal normal eigenlocus  $\tau = \tau_1 - \tau_2$ .



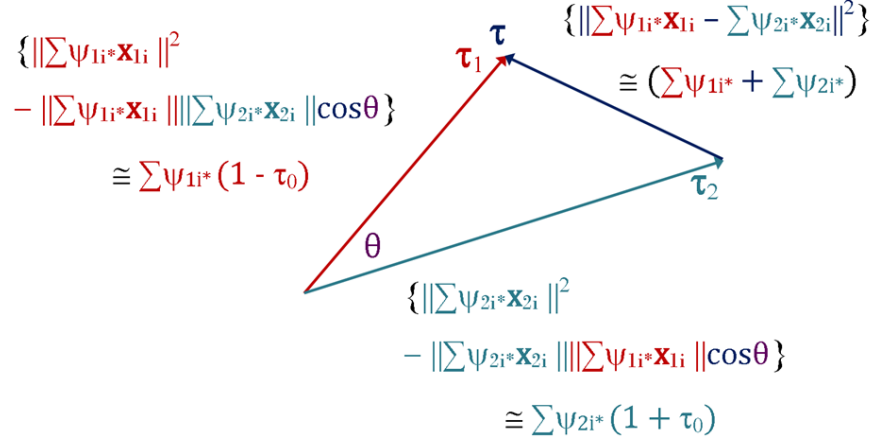


Figure 31: Illustration of how the total allowed eigenenergies of the constrained primal normal eigenaxis components on  $\tau$ ,  $\tau_1$ , and  $\tau_2$  are determined by joint symmetrical distributions of the eigenenergies of  $\psi$  and  $\tau$ . The joint eigenenergies of  $\psi$  and  $\tau$  are symmetrically distributed over the constrained primal normal eigenaxis components on  $\tau = \tau_1 - \tau_2$ .

Properties of the symmetric equalizer statistic  $\nabla_{eq}$  are examined next.

### Geometric and Statistical Properties of the Symmetric Equalizer Statistic $\nabla_{eq}$

The statistical expression for the symmetric equalizer statistic  $\nabla_{eq}$

$$\begin{aligned}
 \nabla_{eq} &= \frac{\tau_0}{2} \sum_{i=1}^l \psi_{i*}, \\
 &= \frac{\tau_0}{2} \|\tau\|_{\min_c}^2, \\
 &= \frac{1}{2l} \|\tau\|_{\min_c}^2 y_i (1 - \xi_i) \\
 &\quad - \frac{1}{2l} \|\tau\|_{\min_c}^2 \sum_{i=1}^l \mathbf{x}_{i*}^T \sum_{j=1}^{l_1} \psi_{1j*} \mathbf{x}_{1j*} \\
 &\quad + \frac{1}{2l} \|\tau\|_{\min_c}^2 \sum_{i=1}^l \mathbf{x}_{i*}^T \sum_{j=1}^{l_2} \psi_{2j*} \mathbf{x}_{2j*},
 \end{aligned}$$

substantiates the significant and joint roles that the KKT constraint of Eq. (38) and the  $\tau_0$  term of Eq. (77) have in achieving the state of statistical equilibrium that is exhibited by a strong dual normal eigenlocus  $\tau = \tau_1 - \tau_2$ .

Indeed, half of the total allowed eigenenergy  $\frac{1}{2} \|\tau\|_{\min_c}^2$  of a strong dual normal eigenlocus is described by integrated lengths of Wolfe dual normal eigenaxis components correlated with each pattern category:

$$\frac{1}{2} \|\tau\|_{\min_c}^2 = \sum_{i=1}^{l_1} \psi_{1i*} = \sum_{i=1}^{l_2} \psi_{2i*},$$

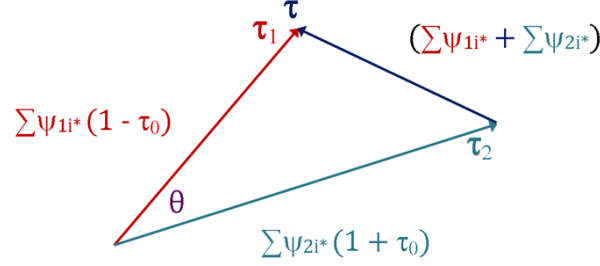
where each Wolfe dual normal eigenaxis component exhibits a critical length which describes the distribution of an extreme training point. Figure 29 illustrates how the KKT constraint of Eq. (38) determines a center of eigenenergy  $\frac{1}{2} \sum_{i=1}^l \psi_{i*}$  for a strong dual normal eigenlocus which is satisfied by half of its total allowed eigenenergy  $\frac{1}{2} \|\tau\|_{\min}^2$ . Figure 30 illustrates how the KKT constraint of Eq. (38) and the  $\tau_0$  term of Eq. (77) jointly ensure that the state of statistical equilibrium exhibited by  $\tau = \tau_1 - \tau_2$  is determined by symmetrically balanced eigenenergies of  $\tau_1$  and  $\tau_2$ .

Equation (77) indicates that the  $\tau_0$  term describes eigen-balanced correlations between the extreme training points and the eigen-balanced eigenloci of the constrained primal normal eigenaxis components on  $\tau_1$  and  $\tau_2$ , where the lengths of the Wolfe dual normal eigenaxis components and the KKT constraint of Eq. (38) have significant and joint roles in balancing highly interconnected sets of inner product relationships amongst the Wolfe dual and the constrained primal normal eigenaxis components. Figure 29 illustrates how the  $\tau_0$  term ensures that the state of statistical equilibrium exhibited by  $\tau = \tau_1 - \tau_2$  is determined by equal and opposite eigenenergies of  $\tau$ ,  $\tau_1$ , and  $\tau_2$ , where the joint eigenenergies of  $\tau$  and  $\psi$  are symmetrically distributed over  $\tau = \tau_1 - \tau_2$ .

The expression for the symmetric equalizer statistic  $\nabla_{eq}$  substantiates the conclusion that effective and consistent fits of constrained primal normal eigenaxis components to training data requires satisfying joint symmetrical distributions of eigenenergies over a constrained primal normal eigenlocus of eigen-scaled extreme data points.

Figure 32 illustrates the joint roles that the KKT constraint of Eq. (38) and the  $\tau_0$  term of Eq. (77) have in achieving the surprisingly elegant statistical balancing feat that is routinely accomplished by solving the inequality constrained optimization problem of Eq. (13). Given the data-driven symmetrical essence of this statistical balancing feat, it is recommended that strong dual normal eigenlocus transforms be applied to equal numbers of training examples from each pattern class.

## A Surprisingly Elegant Statistical Balancing Feat



$$\begin{aligned} \tau_0 = & \frac{1}{l} \sum_{i=1}^l y_i (1 - \xi_i) \\ & - \frac{1}{l} \sum_{i=1}^l \mathbf{x}_{i*}^T \sum_{j=1}^{l_1} \psi_{1j*} \mathbf{x}_{1j*} \\ & + \frac{1}{l} \sum_{i=1}^l \mathbf{x}_{i*}^T \sum_{j=1}^{l_2} \psi_{2j*} \mathbf{x}_{2j*} \end{aligned}$$

Figure 32: Illustration of how symmetrical integrated sets  $\sum_{i=1}^{l_1} \psi_{1i*} \equiv \sum_{i=1}^{l_2} \psi_{2i*}$  of eigen-balanced magnitudes of Wolfe dual normal eigenaxis components, both of which are symmetrically balanced by means of the  $\tau_0$  statistic, jointly ensure that the total allowed eigenenergies of  $\tau_1$  and  $\tau_2$  are symmetrically balanced with each other.

The next section of the paper will examine probabilistic properties which are exhibited by strong dual normal eigenlocus discriminant functions  $D(\mathbf{x}) = \tau^T \mathbf{x} + \tau_0$ . An expression will be obtained for the normal eigenlocus decision rule in Eq. (36) which encodes likelihood ratios. Expressions for the likelihood ratio and the state of statistical equilibrium will be used to show that constrained normal eigenlocus discriminant functions describe linear decision boundaries for which class probabilities are equivalent to each other. The likelihood ratio expression will also be used to reexamine how width regulation of large covariance decision regions is accomplished. It will also be shown that strong dual normal eigenlocus discriminant functions encode Bayes' likelihood ratio for common covariance data and a robust likelihood ratio for all other data distributions.

## 18 Probabilistic Properties Exhibited by the Statistical Equilibrium Point of a Strong Dual Normal Eigenlocus

It will now be shown how strong dual discriminant functions  $D(\mathbf{x}) = \tau^T \mathbf{x} + \tau_0$  encode likelihood ratios. Recall that each constrained primal normal eigenaxis

component  $\psi_{1i_*} \mathbf{x}_{1i_*}$  on  $\tau_1$ , and each constrained primal normal eigenaxis component  $\psi_{2i_*} \mathbf{x}_{2i_*}$  on  $\tau_2$ , provides a maximum covariance estimate in a principal location, in the form of an eigen-balanced first and second order statistical moment about the geometric locus of an extreme data point  $\mathbf{x}_{1i_*}$  or  $\mathbf{x}_{2i_*}$ . It has been demonstrated that any given maximum covariance estimate provides a measure of how much two groups of eigen-scaled extreme data points and their common means vary from a given extreme data point. It has also been demonstrated that any given maximum covariance estimate encodes a distribution of first order coordinates for an extreme training vector  $\mathbf{x}_{1i_*}$  or  $\mathbf{x}_{2i_*}$ , relative to the eigen-scaled extreme vectors for two given data sets. Thereby, any given maximum covariance estimate describes how the components of an extreme training vector are distributed within a collection of eigen-scaled extreme training vectors. A probabilistic explanation for the total allowed eigenenergies of the constrained primal normal eigenaxis components on  $\tau$  is now obtained.

### 18.1 A Probabilistic Explanation for the Total Allowed Eigenenergies of $\tau$

Consider a strong dual normal eigenlocus of constrained primal normal eigenaxis components  $\tau = \sum_{i=1}^{l_1} \psi_{1i_*} \mathbf{x}_{1i_*} - \sum_{i=1}^{l_2} \psi_{2i_*} \mathbf{x}_{2i_*}$  and take any constrained primal normal eigenaxis component  $\psi_{1i_*} \mathbf{x}_{1i_*}$  on  $\tau_1$ . Given that  $\psi_{1i_*} \mathbf{x}_{1i_*}$  provides a maximum covariance estimate in a principal location in  $\mathbb{R}^d$ , it follows that the square  $\|\psi_{1i_*} \mathbf{x}_{1i_*}\|_{\min_c}^2$  of  $\psi_{1i_*} \mathbf{x}_{1i_*}$  is the probability of finding the extreme data point  $\mathbf{x}_{1i_*}$  in a particular region of  $\mathbb{R}^d$ , where the integration  $\|\psi_{1i_*} \mathbf{x}_{1i_*}\|_{\min_c}^2$  of the squared vector components of  $\psi_{1i_*} \mathbf{x}_{1i_*}$  is the total allowed eigenenergy of the constrained primal normal eigenaxis component  $\psi_{1i_*} \mathbf{x}_{1i_*}$ .

Now take any constrained primal normal eigenaxis component  $\psi_{2i_*} \mathbf{x}_{2i_*}$  on  $\tau_2$ . Given that  $\psi_{2i_*} \mathbf{x}_{2i_*}$  provides a maximum covariance estimate in a principal location in  $\mathbb{R}^d$ , it follows that the square  $\|\psi_{2i_*} \mathbf{x}_{2i_*}\|_{\min_c}^2$  of  $\psi_{2i_*} \mathbf{x}_{2i_*}$  is the probability of finding the extreme data point  $\mathbf{x}_{2i_*}$  in a particular region of  $\mathbb{R}^d$ , where the integration  $\|\psi_{2i_*} \mathbf{x}_{2i_*}\|_{\min_c}^2$  of the squared vector components of  $\psi_{2i_*} \mathbf{x}_{2i_*}$  is the total allowed eigenenergy of the constrained primal normal eigenaxis component  $\psi_{2i_*} \mathbf{x}_{2i_*}$ .

It follows that the total allowed eigenenergy  $\|\tau\|_{\min_c}^2$  of  $\tau$

$$\begin{aligned} \|\tau\|_{\min_c}^2 &\cong \left\| \sum_{i=1}^{l_1} \psi_{1i_*} \mathbf{x}_{1i_*} - \sum_{i=1}^{l_2} \psi_{2i_*} \mathbf{x}_{2i_*} \right\|_{\min_c}^2, \\ &\cong \left\| \sum_{i=1}^{l_1} \psi_{1i_*} \mathbf{x}_{1i_*} \right\|_{\min_c}^2 + \left\| \sum_{i=1}^{l_2} \psi_{2i_*} \mathbf{x}_{2i_*} \right\|_{\min_c}^2 \\ &\quad - 2 \left\| \sum_{i=1}^{l_1} \psi_{1i_*} \mathbf{x}_{1i_*} \right\| \left\| \sum_{i=1}^{l_2} \psi_{2i_*} \mathbf{x}_{2i_*} \right\| \cos \theta_{\tau_1 \tau_2}, \end{aligned}$$

describes the probability of finding extreme data points in particular regions of  $\mathbb{R}^d$ , which determines the probability of finding data points in regions of large covariance between either overlapping or non-overlapping data distributions.

It will now be demonstrated that  $\tau$  encodes Bayes' likelihood ratio for common covariance data distributions and a robust likelihood ratio for all other data distributions. An expression is first obtained for  $\tau$  which encodes likelihood ratios.

## 18.2 Likelihood Ratios Encoded Within $\tau$

Returning to the expression for  $\tau$

$$\tau = \sum_{i=1}^{l_1} \psi_{1_{i*}} \mathbf{x}_{1_{i*}} - \sum_{i=1}^{l_2} \psi_{2_{i*}} \mathbf{x}_{2_{i*}},$$

in Eq. (30), and substituting the expressions for  $\psi_{1_{i*}}$  and  $\psi_{2_{i*}}$  in Eqs (61) and (69) into the above expression for  $\tau$ , provides an expression for  $\tau$

$$\begin{aligned} \tau = & \lambda_{\max\psi}^{-1} \sum_{i=1}^{l_1} \frac{\mathbf{x}_{1_{i*}}}{\|\mathbf{x}_{1_{i*}}\|} \|\mathbf{x}_{1_{i*}}\|^2 \times \\ & \left[ \begin{aligned} & \sum_{j=1}^{l_1} \psi_{1_{j*}} \|\mathbf{x}_{1_{j*}}\| \cos \theta_{\mathbf{x}_{1_{i*}} \mathbf{x}_{1_{j*}}} \\ & - \sum_{j=1}^{l_2} \psi_{2_{j*}} \|\mathbf{x}_{2_{j*}}\| \cos \theta_{\mathbf{x}_{1_{i*}} \mathbf{x}_{2_{j*}}} \end{aligned} \right] \\ & - \lambda_{\max\psi}^{-1} \sum_{i=1}^{l_2} \frac{\mathbf{x}_{2_{i*}}}{\|\mathbf{x}_{2_{i*}}\|} \|\mathbf{x}_{2_{i*}}\|^2 \times \\ & \left[ \begin{aligned} & \sum_{j=1}^{l_2} \psi_{2_{j*}} \|\mathbf{x}_{2_{j*}}\| \cos \theta_{\mathbf{x}_{2_{i*}} \mathbf{x}_{2_{j*}}} \\ & - \sum_{j=1}^{l_1} \psi_{1_{j*}} \|\mathbf{x}_{1_{j*}}\| \cos \theta_{\mathbf{x}_{2_{i*}} \mathbf{x}_{1_{j*}}} \end{aligned} \right], \end{aligned} \quad (89)$$

which encodes likelihood ratios, where the terms  $\frac{\|\mathbf{x}_{1_{i*}}\|}{\|\mathbf{x}_{1_{i*}}\|}$  and  $\frac{\|\mathbf{x}_{2_{i*}}\|}{\|\mathbf{x}_{2_{i*}}\|}$  have been introduced and rearranged.

Recall that the eigenloci and corresponding lengths  $\psi_{1_{i*}}$  or  $\psi_{2_{i*}}$  of the Wolfe dual normal eigenaxis components encode pointwise covariance estimates, i.e., maximum covariance estimates in principal locations, for the extreme vectors  $\mathbf{x}_{1_{i*}}$  or  $\mathbf{x}_{2_{i*}}$ . Given Eqs (64) and (72), it follows that Eq. (89) encodes pointwise covariance estimates in the form of eigen-balanced, signed magnitudes along the axes of extreme vectors, where eigen-balanced, signed magnitudes along the axes of the  $\mathbf{x}_{1_{i*}}$  extreme vectors are denoted by

$$\left[ \begin{aligned} & \sum_{j=1}^{l_1} \psi_{1_{j*}} \|\mathbf{x}_{1_{j*}}\| \cos \theta_{\mathbf{x}_{1_{i*}} \mathbf{x}_{1_{j*}}} \\ & - \sum_{j=1}^{l_2} \psi_{2_{j*}} \|\mathbf{x}_{2_{j*}}\| \cos \theta_{\mathbf{x}_{1_{i*}} \mathbf{x}_{2_{j*}}} \end{aligned} \right],$$

and eigen-balanced, signed magnitudes along the axes of the  $\mathbf{x}_{2_{i*}}$  extreme vectors are denoted by

$$\left[ \begin{aligned} & \sum_{j=1}^{l_2} \psi_{2_{j*}} \|\mathbf{x}_{2_{j*}}\| \cos \theta_{\mathbf{x}_{2_{i*}} \mathbf{x}_{2_{j*}}} \\ & - \sum_{j=1}^{l_1} \psi_{1_{j*}} \|\mathbf{x}_{1_{j*}}\| \cos \theta_{\mathbf{x}_{2_{i*}} \mathbf{x}_{1_{j*}}} \end{aligned} \right].$$

Denote the pointwise covariance estimates for the extreme vectors  $\mathbf{x}_{1_{i*}}$  and  $\mathbf{x}_{2_{i*}}$  in Eq. (89) by  $\widehat{\text{cov}}_{up\downarrow}(\mathbf{x}_{1_{i*}})$  and  $\widehat{\text{cov}}_{up\downarrow}(\mathbf{x}_{2_{i*}})$  respectively. Using this notation,

the expression for  $\tau$  in Eq. (89) can be rewritten in the following manner

$$\begin{aligned} \tau = & \lambda_{\max\psi}^{-1} \sum_{i=1}^{l_1} \left\| \sqrt{\widehat{\text{cov}}_{up\uparrow}}(\mathbf{x}_{1i_*}) \mathbf{x}_{1i_*} \right\|_{\min_c}^2 \frac{\mathbf{x}_{1i_*}}{\|\mathbf{x}_{1i_*}\|} \\ & - \lambda_{\max\psi}^{-1} \sum_{i=1}^{l_2} \left\| \sqrt{\widehat{\text{cov}}_{up\uparrow}}(\mathbf{x}_{2i_*}) \mathbf{x}_{2i_*} \right\|_{\min_c}^2 \frac{\mathbf{x}_{2i_*}}{\|\mathbf{x}_{2i_*}\|}. \end{aligned} \quad (90)$$

It follows that each constrained primal normal eigenaxis component  $\psi_{1i_*} \mathbf{x}_{1i_*}$  on  $\tau_1$

$$\psi_{1i_*} \mathbf{x}_{1i_*} \triangleq \lambda_{\max\psi}^{-1} \left\| \sqrt{\widehat{\text{cov}}_{up\uparrow}}(\mathbf{x}_{1i_*}) \mathbf{x}_{1i_*} \right\|_{\min_c}^2 \frac{\mathbf{x}_{1i_*}}{\|\mathbf{x}_{1i_*}\|},$$

describes the probability  $\lambda_{\max\psi}^{-1} \left\| \sqrt{\widehat{\text{cov}}_{up\uparrow}}(\mathbf{x}_{1i_*}) \mathbf{x}_{1i_*} \right\|_{\min_c}^2$  of finding an extreme data point  $\mathbf{x}_{1i_*}$  in a particular region of  $\mathbb{R}^d$ . Thereby, the integrated set of constrained primal normal eigenaxis components on  $\tau_1$

$$\tau_1 = \lambda_{\max\psi}^{-1} \sum_{i=1}^{l_1} \left\| \sqrt{\widehat{\text{cov}}_{up\uparrow}}(\mathbf{x}_{1i_*}) \mathbf{x}_{1i_*} \right\|_{\min_c}^2 \frac{\mathbf{x}_{1i_*}}{\|\mathbf{x}_{1i_*}\|}, \quad (91)$$

describes the probabilities of finding all of the  $\mathbf{x}_{1i_*}$  extreme data points in particular regions of  $\mathbb{R}^d$ , where all of the extreme data points  $\mathbf{x}_{1i_*}$  are located in regions of large covariance between either overlapping or non-overlapping data distributions.

It also follows that each constrained primal normal eigenaxis component  $\psi_{2i_*} \mathbf{x}_{2i_*}$  on  $\tau_2$

$$\psi_{2i_*} \mathbf{x}_{2i_*} \triangleq \lambda_{\max\psi}^{-1} \left\| \sqrt{\widehat{\text{cov}}_{up\uparrow}}(\mathbf{x}_{2i_*}) \mathbf{x}_{2i_*} \right\|_{\min_c}^2 \frac{\mathbf{x}_{2i_*}}{\|\mathbf{x}_{2i_*}\|},$$

describes the probability  $\lambda_{\max\psi}^{-1} \left\| \sqrt{\widehat{\text{cov}}_{up\uparrow}}(\mathbf{x}_{2i_*}) \mathbf{x}_{2i_*} \right\|_{\min_c}^2$  of finding an extreme data point  $\mathbf{x}_{2i_*}$  in a particular region of  $\mathbb{R}^d$ . Thereby, the integrated set of constrained primal normal eigenaxis components on  $\tau_2$

$$\tau_2 = \lambda_{\max\psi}^{-1} \sum_{i=1}^{l_2} \left\| \sqrt{\widehat{\text{cov}}_{up\uparrow}}(\mathbf{x}_{2i_*}) \mathbf{x}_{2i_*} \right\|_{\min_c}^2 \frac{\mathbf{x}_{2i_*}}{\|\mathbf{x}_{2i_*}\|}, \quad (92)$$

describes the probabilities of finding all of the  $\mathbf{x}_{2i_*}$  extreme data points in particular regions of  $\mathbb{R}^d$ , where all of the extreme data points  $\mathbf{x}_{2i_*}$  are located in regions of large covariance between either overlapping or non-overlapping data distributions.

It is concluded that the integrated set of constrained primal normal eigenaxis components on  $\tau_1$  and  $\tau_2$  in Eq. (90) describes the probabilities of finding each of the extreme data points in particular regions of  $\mathbb{R}^d$ , where all of the extreme data points are located in regions of large covariance between either overlapping

or non-overlapping data distributions. Thereby, it is concluded that Eq. (90) *encodes robust and data-driven likelihood ratios*.

It will now be shown that constrained normal eigenlocus discriminant functions describe linear decision boundaries for which class probabilities are equivalent to each other.

### 18.2.1 Equivalence of Class Probabilities

The total allowed eigenenergies  $\|\tau_1 - \tau_2\|_{\min_c}^2$  of the constrained primal normal eigenlocus components  $\tau_1$  and  $\tau_2$  on  $\tau$  satisfy a linear decision boundary and the bilaterally symmetrical borders which bound it. Returning to Eq. (87), it is concluded that the state of statistical equilibrium which is satisfied by the total allowed eigenenergies of  $\tau_1$  and  $\tau_2$

$$E_{\tau_1} + \frac{\tau_0}{2} \|\tau\|_{\min_c}^2 \Leftrightarrow E_{\tau_2} - \frac{\tau_0}{2} \|\tau\|_{\min_c}^2,$$

in relation to the centrally located statistical fulcrum  $f_s$

$$f_s = \frac{1}{2} \|\tau\|_{\min_c}^2,$$

ensures that *the integrated sum of probabilities encoded with the normal eigenaxis components on  $\tau_1$ :*

$$\lambda_{\max_\psi}^{-1} \sum_{i=1}^{l_1} \left\| \sqrt{\widehat{\text{COV}}_{up\downarrow}}(\mathbf{x}_{1i*}) \mathbf{x}_{1i*} \right\|_{\min_c}^2 \frac{\mathbf{x}_{1i*}}{\|\mathbf{x}_{1i*}\|},$$

*are balanced with the integrated sum of probabilities encoded with the normal eigenaxis components on  $\tau_2$ :*

$$\lambda_{\max_\psi}^{-1} \sum_{i=1}^{l_2} \left\| \sqrt{\widehat{\text{COV}}_{up\downarrow}}(\mathbf{x}_{2i*}) \mathbf{x}_{2i*} \right\|_{\min_c}^2 \frac{\mathbf{x}_{2i*}}{\|\mathbf{x}_{2i*}\|}.$$

It follows that the probabilities described by Eq. (91) are equivalent to the probabilities described by Eq. (92). Thereby, the probabilities  $P(\mathbf{X}_1)$  of the pattern class  $\mathbf{X}_1$  are equivalent to the probabilities  $P(\mathbf{X}_2)$  of the pattern class  $\mathbf{X}_2$

$$P(\mathbf{X}_1) \equiv P(\mathbf{X}_2). \quad (93)$$

It is concluded that the statistical equilibrium point of  $\tau$  determines large covariance decision regions, i.e., linear decision boundaries and regulated linear decision borders, for which class probabilities are equivalent to each other.

Recall that the geometric loci of a linear decision boundary  $D_0(\mathbf{x})$  and its bilaterally symmetrical linear decision borders  $D_1(\mathbf{x})$  and  $D_{-1}(\mathbf{x})$  are determined by Eqs (23), (24), and (25). Returning to Eqs (31), (32), and (33), recall that the eigenloci of the constrained primal normal eigenlocus components  $\tau_1$  and  $\tau_2$  regulate the geometric width, i.e., the breadth, of the geometric region between the linear decision borders  $D_1(\mathbf{x})$  and  $D_{-1}(\mathbf{x})$ . The eigenloci of  $\tau_1$  and

$\tau_2$  also regulate the span of the congruent geometric regions between the linear decision boundary  $D_0(\mathbf{x})$  and the linear decision borders  $D_{+1}(\mathbf{x})$  and  $D_{-1}(\mathbf{x})$ . The next part of the paper will reconsider how width regulation of large covariance decision regions is accomplished. The analysis will provide probabilistic explanations for how Eq. (31) describes geometric regions of large covariance between non-overlapping data distributions and overlapping data distributions. The analysis will also provide probabilistic explanations for how Eqs (32) and (33) describe disjoint tail regions between non-overlapping data distributions, and bipartite, joint geometric regions of large covariance between overlapping data distributions.

### 18.3 Probabilistic Expressions of Decision Region Widths

Consider again Eq. (31):

$$D_{(D_1(\mathbf{x})-D_{-1}(\mathbf{x}))} = \frac{2}{\|\tau_1 - \tau_2\|},$$

which describes the width of the geometric region of large covariance between the linear decision borders  $D_1(\mathbf{x})$  and  $D_{-1}(\mathbf{x})$ . Substituting the expressions for  $\tau_1$  in Eq. (91) and  $\tau_2$  in Eq. (92) into Eq. (31) provides a probabilistic expression for the span of the geometric region of large covariance between the linear decision borders  $D_1(\mathbf{x})$  and  $D_{-1}(\mathbf{x})$ :

$$D_{(D_1(\mathbf{x})-D_{-1}(\mathbf{x}))} = 2 \left\| \begin{array}{l} \lambda_{\max_\psi}^{-1} \sum_{i=1}^{l_1} \left\| \sqrt{\widehat{\text{cov}}_{up_\uparrow}}(\mathbf{x}_{1_{i*}}) \mathbf{x}_{1_{i*}} \right\|_{\min_c}^2 \frac{\mathbf{x}_{1_{i*}}}{\|\mathbf{x}_{1_{i*}}\|} \\ - \lambda_{\max_\psi}^{-1} \sum_{i=1}^{l_2} \left\| \sqrt{\widehat{\text{cov}}_{up_\uparrow}}(\mathbf{x}_{2_{i*}}) \mathbf{x}_{2_{i*}} \right\|_{\min_c}^2 \frac{\mathbf{x}_{2_{i*}}}{\|\mathbf{x}_{2_{i*}}\|} \end{array} \right\|^{-1}.$$

Given the above expression, it is concluded that the width of the geometric region of large covariance between the linear decision borders  $D_1(\mathbf{x})$  and  $D_{-1}(\mathbf{x})$  is regulated by probabilities that extreme data points are located in particular regions of  $\mathbb{R}^d$ .

Now recall that the span of the congruent regions between the linear decision boundary  $D_0(\mathbf{x})$  and the linear decision borders  $D_{+1}(\mathbf{x})$  and  $D_{-1}(\mathbf{x})$  is regulated by the expression

$$\frac{1}{\|\tau_1 - \tau_2\|}.$$

Substituting the expressions for  $\tau_1$  in Eq. (91) and  $\tau_2$  in Eq. (92) into Eqs (32) and (33), provides probabilistic expressions for the span of the congruent geometric regions between the linear decision boundary  $D_0(\mathbf{x})$  and the linear decision borders  $D_{+1}(\mathbf{x})$  and  $D_{-1}(\mathbf{x})$ , where the width of the geometric region of large covariance between the linear decision boundary  $D_0(\mathbf{x})$  and the linear decision border  $D_{+1}(\mathbf{x})$  described by the expression:

$$D_{(D_0(\mathbf{x})-D_{+1}(\mathbf{x}))} = \left\| \begin{array}{l} \lambda_{\max_\psi}^{-1} \sum_{i=1}^{l_1} \left\| \sqrt{\widehat{\text{cov}}_{up_\uparrow}}(\mathbf{x}_{1_{i*}}) \mathbf{x}_{1_{i*}} \right\|_{\min_c}^2 \frac{\mathbf{x}_{1_{i*}}}{\|\mathbf{x}_{1_{i*}}\|} \\ - \lambda_{\max_\psi}^{-1} \sum_{i=1}^{l_2} \left\| \sqrt{\widehat{\text{cov}}_{up_\uparrow}}(\mathbf{x}_{2_{i*}}) \mathbf{x}_{2_{i*}} \right\|_{\min_c}^2 \frac{\mathbf{x}_{2_{i*}}}{\|\mathbf{x}_{2_{i*}}\|} \end{array} \right\|^{-1},$$



and the width of the geometric region of large covariance between the linear decision boundary  $D_0(\mathbf{x})$  and the linear decision border  $D_{-1}(\mathbf{x})$  described by the equivalent expression:

$$D(D_0(\mathbf{x})-D_{-1}(\mathbf{x})) = \left\| \begin{array}{c} \lambda_{\max_\psi}^{-1} \sum_{i=1}^{l_1} \left\| \sqrt{\widehat{\text{COV}}_{up_\dagger}(\mathbf{x}_{1_{i*}})} \mathbf{x}_{1_{i*}} \right\|_{\min_c}^2 \frac{\mathbf{x}_{1_{i*}}}{\|\mathbf{x}_{1_{i*}}\|} \\ - \lambda_{\max_\psi}^{-1} \sum_{i=1}^{l_2} \left\| \sqrt{\widehat{\text{COV}}_{up_\dagger}(\mathbf{x}_{2_{i*}})} \mathbf{x}_{2_{i*}} \right\|_{\min_c}^2 \frac{\mathbf{x}_{2_{i*}}}{\|\mathbf{x}_{2_{i*}}\|} \end{array} \right\|^{-1},$$

is regulated by probabilities of finding extreme data points within particular regions of  $\mathbb{R}^d$ .

Given the above expressions, it is concluded that the geometric loci of a linear decision boundary and its bilaterally symmetrical linear decision borders, which are determined by Eqs (23), (24), and (25), are regulated by probabilities of finding extreme data points within particular regions of  $\mathbb{R}^d$ .

Given all of the above expressions and Eq. (93), it is also concluded that the geometric loci of a linear decision boundary and its bilaterally symmetrical linear decision borders, which are determined by Eqs (23), (24), and (25), are regulated by class probabilities that are equivalent to each other.

## Equivalence Between Bayes' Likelihood Ratio and the Normal Eigenlocus Likelihood Ratio

It will now be shown that the normal eigenlocus test statistic in Eq. (36) encodes the Bayes' likelihood ratio expression for similar covariance data distributions. An expression for the normal eigenlocus likelihood ratio is now obtained.

### 18.4 The Normal Eigenlocus Likelihood Ratio

Consider again the normal eigenlocus test statistic  $\Lambda_\tau(\mathbf{x}) \underset{H_2}{\overset{H_1}{\geq}} 0$  in Eq. (36)

$$\begin{aligned} \Lambda_\tau(\mathbf{x}) &= \left( \mathbf{x} - \frac{1}{l} \sum_{i=1}^l \mathbf{x}_{i*} \right)^T \tau_1 \\ &\quad - \left( \mathbf{x} - \frac{1}{l} \sum_{i=1}^l \mathbf{x}_{i*} \right)^T \tau_2 \\ &\quad + \frac{1}{l} \sum_{i=1}^l y_i (1 - \xi_i) \underset{H_2}{\overset{H_1}{\geq}} 0, \end{aligned}$$

where  $\tau = \tau_1 - \tau_2$ . Substituting the expressions for  $\tau_1$  and  $\tau_2$  in Eqs (91) and (92) into the above expression for  $\Lambda_\tau(\mathbf{x})$  provides an expression for the normal

eigenlocus decision rule  $\Lambda_{\tau_1-\tau_2}(\mathbf{x})$  in terms of likelihoods

$$\begin{aligned} \Lambda_{\tau_1-\tau_2}(\mathbf{x}) = & \left( \mathbf{x} - \frac{1}{l} \sum_{i=1}^l \mathbf{x}_{i*} \right)^T \times \\ & \left[ \lambda_{\max_\psi}^{-1} \sum_{i=1}^{l_1} \left\| \sqrt{\text{cov}_{up\downarrow}}(\mathbf{x}_{1i*}) \mathbf{x}_{1i*} \right\|_{\min_c}^2 \frac{\mathbf{x}_{1i*}}{\|\mathbf{x}_{1i*}\|} \right] \\ & - \left( \mathbf{x} - \frac{1}{l} \sum_{i=1}^l \mathbf{x}_{i*} \right)^T \times \\ & \left[ \lambda_{\max_\psi}^{-1} \sum_{i=1}^{l_2} \left\| \sqrt{\text{cov}_{up\downarrow}}(\mathbf{x}_{2i*}) \mathbf{x}_{2i*} \right\|_{\min_c}^2 \frac{\mathbf{x}_{2i*}}{\|\mathbf{x}_{2i*}\|} \right] \\ & + \frac{1}{l} \sum_{i=1}^l y_i (1 - \xi_i) \underset{H_2}{\overset{H_1}{\geq}} 0, \end{aligned} \quad (94)$$

where the terms  $\lambda_{\max_\psi}^{-1} \left\| \sqrt{\text{cov}_{up\downarrow}}(\mathbf{x}_{1i*}) \mathbf{x}_{1i*} \right\|_{\min_c}^2$  and  $\lambda_{\max_\psi}^{-1} \left\| \sqrt{\text{cov}_{up\downarrow}}(\mathbf{x}_{2i*}) \mathbf{x}_{2i*} \right\|_{\min_c}^2$  describe the likelihood of finding an extreme data point in a particular region of  $\mathbb{R}^d$ . The normal eigenlocus ratio test in Eq. (94) is now compared with Bayes' decision rule for common covariance data.

### 18.5 Comparison of the Normal Eigenlocus Decision Rule with Bayes' Decision Rule

Bayes' decision rule and boundary are completely defined within the likelihood ratio expression  $\Lambda(\mathbf{x})$ :

$$\begin{aligned} \Lambda(\mathbf{x}) = & \frac{|\Sigma_2|^{1/2} \exp \left\{ -\frac{1}{2} (\mathbf{x} - \mu_1)^T \Sigma_1^{-1} (\mathbf{x} - \mu_1) \right\}}{|\Sigma_1|^{1/2} \exp \left\{ -\frac{1}{2} (\mathbf{x} - \mu_2)^T \Sigma_2^{-1} (\mathbf{x} - \mu_2) \right\}} \\ & \underset{H_2}{\overset{H_1}{\geq}} \frac{P_2 (C_{12} - C_{22})}{P_1 (C_{21} - C_{11})}, \end{aligned}$$

for Gaussian data, where no costs ( $C_{11} = C_{22} = 0$ ) are associated with correct decisions and  $C_{12} = C_{21} = 1$  are the costs associated with incorrect decisions Duda et al. [2001], VanTrees [1968]. The probabilistic expression of  $\Lambda(\mathbf{x})$  can be reduced to an algebraic expression by means of a natural logarithm transform  $\ln[\Lambda(\mathbf{x})]$ .

The natural logarithm of the likelihood ratio  $\ln[\Lambda(\mathbf{x})]$ :

$$\begin{aligned} \ln[\Lambda(\mathbf{x})] = & \mathbf{x}^T (\Sigma_1^{-1} \mu_1 - \Sigma_2^{-1} \mu_2) \\ & + \frac{1}{2} \mathbf{x}^T (\Sigma_2^{-1} \mathbf{x} - \Sigma_1^{-1} \mathbf{x}) \\ & + \frac{1}{2} \mu_2^T \Sigma_2^{-1} \mu_2 - \frac{1}{2} \mu_1^T \Sigma_1^{-1} \mu_1 \underset{H_2}{\overset{H_1}{\geq}} \eta, \end{aligned} \quad (95)$$

$$\eta \triangleq \ln(P_2) - \ln(P_1) + \frac{1}{2} \ln(|\mathbf{\Sigma}_1|) - \frac{1}{2} \ln(|\mathbf{\Sigma}_2|),$$

produces an algebraic expression that defines the general form of the discriminant function for the general Gaussian binary classification problem, where  $\mu_1$  and  $\mu_2$  are  $d$ -component mean vectors,  $\mathbf{\Sigma}_1$  and  $\mathbf{\Sigma}_2$  are  $d$ -by- $d$  covariance matrices,  $\mathbf{\Sigma}^{-1}$  and  $|\mathbf{\Sigma}|$  denote the inverse and determinant of a covariance matrix respectively,  $H_1$  or  $H_2$  is the true data category, and the class probabilities  $P_1 = P_2 = 1/2$ .

Bayesian decision theory provides the result that the algebraic expression in Eq. (95) describes the geometric loci of Bayes' decision boundaries for any two classes of data drawn from Gaussian distributions. Bayes' decision boundaries are defined by regions for which class probabilities are equivalent to each other, i.e.,  $P_1 = P_2 = 1/2$ , and are characterized by the class of hyperquadric decision surfaces which include hyperplanes, pairs of hyperplanes, hyperspheres, hyperellipsoids, hyperparaboloids, and hyperhyperboloids Duda et al. [2001], VanTrees [1968].

### Bayes' Decision Rule for Similar Covariance Data

Letting  $\mathbf{\Sigma}_1^{-1} = \mathbf{\Sigma}_2^{-1} = \mathbf{\Sigma}^{-1}$  in Eq. (95) provides an expression

$$\begin{aligned} \ln[\Lambda(\mathbf{x})] &= \mathbf{x}^T \mathbf{\Sigma}^{-1} (\mu_1 - \mu_2) \\ &+ \frac{1}{2} \mu_2^T \mathbf{\Sigma}^{-1} \mu_2 - \frac{1}{2} \mu_1^T \mathbf{\Sigma}^{-1} \mu_1 \underset{H_2}{\overset{H_1}{\gtrless}} \eta, \end{aligned} \quad (96)$$

$$\eta \triangleq \ln(P_2) - \ln(P_1),$$

which defines the general form of the discriminant function for similar covariance Gaussian data. Bayesian decision theory provides the result that Eq. (96) encodes Bayes' likelihood ratio for similar covariance Gaussian data. Bayesian decision theory also provides the result that the algebraic expression in Eq. (96) describes the geometric loci of Bayes' linear decision boundaries Duda et al. [2001], VanTrees [1968].

Now, consider again the normal eigenlocus likelihood ratio test in Eq. (94). Given that linear kernel SVM learns Bayes' linear decision boundaries, and given that Eq. (96) determines the locus equation of Bayes' linear decision boundaries, it follows that the constrained normal eigenlocus discriminant function  $\Lambda_{\tau_1 - \tau_2}(\mathbf{x})$  in Eq. (94) determines the locus equation of linear decision boundaries for similar covariance Gaussian data. Indeed, it has been shown that the geometric loci of linear decision boundaries determined by the constrained normal eigenlocus discriminant function  $\Lambda_{\tau_1 - \tau_2}(\mathbf{x})$  in Eq. (94) are defined by regions for which class probabilities are equivalent to each other.

It is concluded that the geometric loci of Bayes' linear decision boundaries are completely defined within the normal eigenlocus likelihood ratio expression

$$\Lambda_{\tau_1 - \tau_2}(\mathbf{x}) \underset{H_2}{\overset{H_1}{\gtrless}} 0 \text{ in Eq. (94).}$$

Now, *Bayes' decision rule and boundary are completely defined within the likelihood ratio expression*  $\Lambda(\mathbf{x})$  in Eq. (96). Therefore, given that the constrained normal eigenlocus discriminant function  $\Lambda_\tau(\mathbf{x})$  describes the geometric loci of Bayes' linear decision boundaries, it also follows that the likelihood ratio encoded within the normal eigenlocus test statistic  $\Lambda_{\tau_1-\tau_2}(\mathbf{x}) \underset{H_2}{\overset{H_1}{\geq}} 0$  in Eq. (94) describes Bayes' likelihood ratio for similar covariance Gaussian data. Numerous simulations studies show that the normal eigenlocus decision rule in Eq. (94) achieves Bayes' error rate for normally distributed data that have the same covariance matrix, including homogeneous data distributions.

It is concluded that the likelihood ratio  $\Lambda_{\tau_1-\tau_2}(\mathbf{x}) \underset{H_2}{\overset{H_1}{\geq}} 0$  encoded within the discriminant function  $D(\mathbf{x}) = \tau^T \mathbf{x} + \tau_0$  determines Bayes' likelihood ratio for similar covariance Gaussian data.

Clearly, then, the normal eigenlocus test statistic  $\Lambda_{\tau_1-\tau_2}(\mathbf{x}) \underset{H_2}{\overset{H_1}{\geq}} 0$  in Eq. (94) provides a sufficient statistic for Bayes' decision rule and boundary for similar covariance data. It is concluded that Bayes' decision rule and boundary are completely defined within the normal eigenlocus likelihood ratio expression  $\Lambda_{\tau_1-\tau_2}(\mathbf{x}) \underset{H_2}{\overset{H_1}{\geq}} 0$  in Eq. (94).

Furthermore, given the robust, data-driven likelihood ratio encoded within the normal eigenlocus decision rule  $\Lambda_{\tau_1-\tau_2}(\mathbf{x})$  in Eq. (94), it is concluded that the normal eigenlocus test statistic provides a robust decision statistic for all other data distributions.

The above analysis substantiates the previous work that is outlined next.

## 18.6 Previous Work on Linear Kernel SVMs

It has been demonstrated that  $\Lambda_{\tau_1-\tau_2}(\mathbf{x}) \equiv \ln[\Lambda(\mathbf{x})]$  for the normally distributed training data described next. This class of Gaussian data is considered to be linearly separable, i.e., a separating line, plane, or hyperplane exists for all such Gaussian data.

### Linearly Separable Data

Linear curves and surfaces provide optimal decision boundaries for normally distributed data sets that have the same covariance matrix  $\Sigma_0 = \Sigma_1 = \Sigma$ :

$$\begin{aligned} p(\mathbf{x}|H_0) &\sim N(\mu_0, \Sigma), \\ p(\mathbf{x}|H_1) &\sim N(\mu_1, \Sigma). \end{aligned}$$

This class of problems has been referred to as a linearly separable classification problem Reeves [2009].

### Linearly Separable Classification Problem

Consider the following binary classification problem:

$$\mathbf{x}_i = \begin{cases} \mathbf{s}_0 + \mathbf{n}_i; & H_0 \\ \mathbf{s}_1 + \mathbf{n}_i; & H_1 \end{cases}, \quad (97)$$

where  $\mathbf{x}_i$  is the  $i^{th}$   $d \times 1$  random vector under hypotheses  $H_0$  and  $H_1$ , respectively,  $\mathbf{n}_i$  is the corresponding noise vector, and  $\mathbf{s}_0$  and  $\mathbf{s}_1$  are deterministic signal vectors. It is assumed that the  $\mathbf{n}_i$  are independent and identically distributed (i.i.d.) zero-mean correlated Gaussian random vectors with known  $d \times d$  noise covariance matrix  $R_N$ . Because  $\mathbf{s}_0$  and  $\mathbf{s}_1$  are deterministic and the noise is additive and Gaussian, Eq. (97) defines a linearly separable classification problem. The data points  $\mathbf{x}_i$  are optimally partitioned by linear curves and surfaces, and are said to be linearly separable.

Let  $\tilde{\mathbf{X}} \triangleq \mathbf{D}_y \mathbf{X}$ , where  $\mathbf{D}_y$  is a  $N \times N$  diagonal matrix with of training labels  $y_i$  and the  $N \times d$  data matrix is  $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N)^T$ . It has been shown that a strong dual normal eigenlocus decision test  $D(\mathbf{x}) \gtrless_{H_0}^{H_1} 0$  can be written as:

$$\begin{aligned} \mathbf{m}_1^T \mathbf{R}_N^{-1} \mathbf{x} &\gtrless_{H_0}^{H_1} (1 - p_0) \mathbf{m}_1^T \mathbf{R}_N^{-1} \mathbf{m}_1 \\ &- (1 - 2p_0) \left\{ \frac{1 + p_0 (1 - p_0) \mathbf{m}_1^T \mathbf{R}_N^{-1} \mathbf{m}_1}{2p_0 (1 - p_0) (1 + \bar{\lambda})} \right\}, \end{aligned}$$

where the correlation matrix  $\mathbf{R}_x \doteq N^{-1} \tilde{\mathbf{X}}^T \tilde{\mathbf{X}}$  can be expressed as  $\mathbf{R}_x = \mathbf{R}_N + (1 - p_0) \mathbf{m}_1 \mathbf{m}_1^T$  and  $\mathbf{R}_x^{-1} = (\mathbf{R}_N (\mathbf{I} + (1 - p_0) \mathbf{R}_N^{-1} \mathbf{m}_1 \mathbf{m}_1^T))^{-1}$ , the quantity  $p_0 \mathbf{m}_0 + (1 - p_0) \mathbf{m}_1 \doteq \sum_{i=1}^N \mathbf{x}_i / N$ ,  $\mathbf{m}_0$  and  $\mathbf{m}_1$  are the means under  $H_0$  and  $H_1$  respectively, and  $p_0$  is the probability that the  $H_0$  hypothesis is represented in the training data Reeves [2009]. Without loss of generality,  $\mathbf{m}_0 = 0$ .

The strong dual normal eigenlocus decision test is now compared with Bayes' test.

### Comparison with the Bayes' Test

By way of comparison, recall that the Bayes' Test is:

$$\begin{aligned} \mathbf{m}_1^T \mathbf{R}_N^{-1} \mathbf{x} &\gtrless_{H_0}^{H_1} (1 - p_0) \mathbf{m}_1^T \mathbf{R}_N^{-1} \mathbf{m}_1 \\ &+ \ln \left( \frac{p_0 (C_{10} - C_{00})}{p_1 (C_{01} - C_{11})} \right). \end{aligned}$$

When  $p_0 = 1/2$  and  $C_{10} = C_{01} = 1$  and  $C_{00} = C_{11} = 0$  (no costs are associated with correct decisions), the normal eigenlocus and the Bayes' tests are equivalent. However, when  $p_0 \neq 1/2$ , the hypothesis tests are different. Although the

sufficient statistic  $\mathbf{m}_1^T \mathbf{R}_N^{-1} \mathbf{x}$  is the same, the thresholds are different. This is because the analogous costs associated with the strong dual normal eigenlocus test are functions of  $p_0$ , whereas for the Bayes' test, these costs are fixed. It is concluded that the discriminant function  $D(\mathbf{x})$  in Eq. (22) encodes an optimal test statistic that is the minimum probability of error for making a decision for the training data described by Eq. (97). The above analysis has been validated with simulation studies Reeves [2009].

The next section of the paper will consider dual-use of strong dual normal eigenlocus discriminant functions. Dual-use involves the practical matter of building robust, scalable, and optimal, probabilistic, multiclass, linear pattern recognition systems. Dual-use also involves a statistical multimeter which effectively measures class separability and Bayes' error rate. The statistical multimeter provides a robust indicator of homogeneous data distributions.

## 19 Design of Probabilistic Multiclass Linear Pattern Recognition Systems

Given the robust, data-driven likelihood ratio expression encoded within the strong dual normal eigenlocus discriminant function in Eq. (94), it follows that the constrained discriminant function in Eq. (22) describes robust linear decision boundaries for any given data distributions. Indeed, the data-driven likelihood ratio expression  $\Lambda_{\tau_1 - \tau_2}(\mathbf{x}) \underset{H_2}{\overset{H_1}{\geq}} 0$  encoded within Eq. (94) provides

a robust test statistic for both overlapping and non-overlapping data distributions. Moreover, it has been demonstrated that strong dual normal eigenlocus transforms produce *regularized and customized statistical decision systems* for the binary classification task. Figures 17, 18, 19, 20, and 22 illustrate how constrained normal eigenlocus discriminant functions determine regularized and customized, data-driven geometric architectures that encode robust decision statistics for the binary classification task.

Given the robust, data-driven likelihood ratio expression encoded within the probabilistic linear discriminant function  $D(\mathbf{x}) = \tau^T \mathbf{x} + \tau_0$ , it follows that strong dual normal eigenlocus discriminant functions provide *robust statistical building blocks for probabilistic, multiclass, linear pattern recognition systems*. A strong dual statistical decision function  $\text{sign}(\Lambda_\tau(\mathbf{x}))$

$$\text{sign}(\Lambda_\tau(\mathbf{x})) = \text{sign} \left[ \left( \mathbf{x} - \frac{1}{l} \sum_{i=1}^l \mathbf{x}_{i*} \right)^T \tau + \dots \right] \\ \text{sign} \left[ \dots + \frac{1}{l} \sum_{i=1}^l y_i (1 - \xi_i) \right],$$

where  $\text{sign}(x) \equiv \frac{x}{|x|}$  for  $x \neq 0$ , provides a natural means for discriminating between multiple classes of data, where robust or optimal decisions can be made that are based on the largest probabilistic output of decision banks of strong dual

statistical decision functions  $\text{sign}(\Lambda_\tau(\mathbf{x}))$ . Figure 33 illustrates the structure of a scalable statistical building block of a statistical bank used to build a probabilistic statistical decision engine which distinguishes between objects in  $M$  different pattern classes.

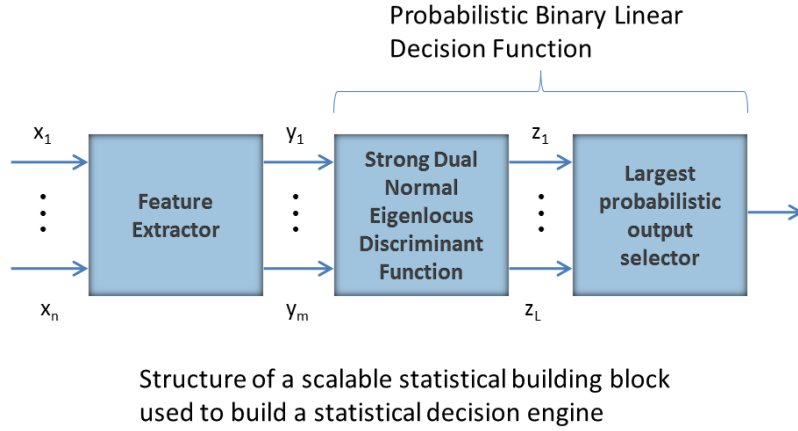


Figure 33: Illustration of a scalable statistical building block of a statistical decision bank  $D_{dbX_i}(\mathbf{x})$  used to build a probabilistic statistical decision engine  $PD_e[\mathbf{x}]$  which distinguishes between objects in  $M$  different pattern classes.

### 19.1 Design of Customized Probabilistic Statistical Decision Engines

Consider the design of a probabilistic, multiclass, linear pattern recognition system that distinguishes between objects in  $M$  different pattern classes. The powerful statistical machinery encoded within the binary class statistical decision function  $\text{sign}(\Lambda_\tau(\mathbf{x}))$  enables the design of a customized, probabilistic statistical decision engine  $PD_e[\text{sign}(\Lambda_\tau(\mathbf{x}))]$  that recognizes the objects in each of the  $M$  pattern classes.

The design of a probabilistic statistical decision engine  $PD_e[\text{sign}(\Lambda_\tau(\mathbf{x}))]$  involves designing  $M \times (M-1)$  strong dual statistical decision functions  $\text{sign}(\Lambda_\tau(\mathbf{x}))$ , each of which consists of a feature extractor and a normal eigenlocus discriminant function. Accordingly, a statistical decision bank  $D_{dbX_i}[\{\text{sign}(\Lambda_{\tau_j}(\mathbf{x}))\}_{j=1}^{M-1}]$  can be developed for each given pattern class  $X_i$  that consists of a bank of  $M-1$  statistical decision functions  $\text{sign}(\Lambda_\tau(\mathbf{x}))$ , where the pattern vectors in the given

class  $X_i$  have the training label +1, and the statistical decision bank  $D_{dbX_i}$  is a linear combination of customized statistical decision functions

$$D_{dbX_i} = \sum_{j=1}^{M-1} \text{sign}(\Lambda_{\tau_j}(\mathbf{x})).$$

The probabilistic statistical decision engine

$$PD_e \left[ \left\{ D_{dbX_i} \left[ \left\{ \text{sign}(\Lambda_{\tau_j}(\mathbf{x})) \right\}_{j=1}^{M-1} \right] \right\}_{i=1}^M \right],$$

provides a set of  $M \times (M - 1)$  decision statistics  $\text{sign}(\Lambda_{\tau}(\mathbf{x}))$ , where the maximum value selector of the statistical decision engine chooses the pattern class  $X_i$  for which a statistical decision bank  $D_{dbX_i}(\mathbf{x})$  has *the maximum probabilistic output*. Because the statistical decision engine  $PD_e[\mathbf{x}]$  is a linear combination of discriminant functions, the overall network complexity is scale-invariant for the feature space dimension and the number of pattern classes.

Furthermore, if a feature extractor has been developed that generates non-overlapping feature vectors for all of the  $M$  pattern classes, then the scale-invariance of the statistical decision engine  $PD_e[\mathbf{x}]$  ensures low estimation variance and optimal generalization performance for feature vectors that possess optimal discrimination capacity. All classes of feature vectors drawn from non-overlapping probability distributions, for which Bayes' error is zero, naturally exhibit optimal discrimination capacity.

The next section considers the use of normal eigenlocus test statistics to design effective feature extractors. Strong dual normal eigenlocus decision functions  $\text{sign}(\Lambda_{\tau}(\mathbf{x}))$  provide a robust statistical multimeter for measuring class separability and Bayes' error rate.

## 19.2 Practical Dual-Use of Strong Dual Normal Eigenlocus Discriminant Functions

Recall that machine learning algorithms for classification systems introduce four sources of error, i.e., Bayes' error, modeling error, estimation error, and computational error, into the final classification system. Bayes' error, i.e., the probability of error, results from overlap between data distributions. Given the robust, stable, and probabilistic properties of strong dual discriminant functions, *the fundamental and difficult problem that remains to be solved is the design of an effective feature extractor*. Given two or more pattern classes, the design of a feature extractor involves determining measurements or features which are most effective for preserving class separability, where class separability is equivalent to the probability of error due to the Bayes' classifier. A feature extractor produces characteristic signatures, called feature vectors, that describe the objects in a pattern class, such as fingerprints or voices. A typical characteristic signature is an ordered sequence of measurements, whereby each measurement describes a numerical attribute or feature of an object in a pattern class. Numerical features are random variables that are characterized by expected values and covariances.



Common examples of characteristic signatures include genetic signatures, proteomic signatures, geometric shape recognition signatures, chemical signatures, spectral signatures, biological signatures, radar signatures, lidar signatures, and multispectral or hyperspectral signatures. This list is by no means exhaustive. Feature vectors can be extracted from any given collection of signals or images.

*The probability of error is the key parameter of all statistical pattern recognition systems.* The amount of overlap between data distributions determines the Bayes' classification error rate which is the best error rate that can be achieved by any classifier Fukunaga [1990].

### 19.3 Using Normal Eigenlocus Test Statistics to Design Effective Feature Extractors

A critical design objective for any statistical pattern recognition system is to develop a feature extractor that provides distinct statistical signatures for all of the pattern classes, i.e., negligible or no overlap exists amongst each pair of data distributions. *Moreover, the criteria to evaluate the effectiveness of features must be a measure of the overlap or class separability among data distributions, and not a measure of fit such as the mean-square error of a statistical model* Fukunaga [1990].

In general, Bayes' error rate is difficult to evaluate. Explicit mathematical expressions are only available for a few special cases. For normal distributions, calculation of the Bayes' error involves numerical integration, except for the common covariance case. Alternatively, the Bhattacharyya distance provides a convenient measure of class separability for two pattern classes. In addition, the Bhattacharyya distance provides an upper bound of the Bayes' error, if training data are drawn from Gaussian distributions. However, the Bhattacharyya distance is difficult to evaluate because the trace and the determinant of matrices are combined in the criterion Fukunaga [1990].

*On the other hand, strong dual normal eigenlocus discriminant functions provide a robust measure of class separability and Bayes' error rate for any given sets of feature vectors.*

#### 19.3.1 A Robust Statistical Multimeter

Strong dual statistical decision functions  $\text{sign}(\Lambda_\tau(\mathbf{x}))$  provide a useful statistical multimeter for measuring data distribution overlap and Bayes' error rate. Given the robust, data-driven likelihood ratio test encoded within strong dual discriminant functions, strong dual statistical decision functions  $\text{sign}(\Lambda_\tau(\mathbf{x}))$  can be used to estimate data distribution overlap and Bayes' error rate for any given sets of feature vectors. In addition, strong dual statistical decision functions  $\text{sign}(\Lambda_\tau(\mathbf{x}))$  can be used to identify homogeneous data distributions. Given any homogeneous data distribution, it has been shown that (1) most, if not all, of the training data are transformed into constrained primal normal eigenaxis

components, and (2) the error rate of the strong dual discriminant function is 50%.

## 19.4 Summary of Practical Dual-Use of Linear Kernel SVMs

It is concluded that strong dual statistical decision functions  $\text{sign}(\Lambda_\tau(\mathbf{x}))$  have practical dual-use as (1) statistical multimeters in the design of effective feature extractors, (2) robust indicators of homogeneous data distributions, and (3) statistical building blocks of statistical decision banks  $D_{dbX_i}(\mathbf{x})$  used to form probabilistic statistical decision engines  $PD_e[\mathbf{x}]$ . *Thereby, it is concluded that linear kernel SVMs are a powerful and robust class of statistical learning machines which are useful for the design, development, and implementation of probabilistic, multiclass linear pattern recognition systems.*

The final sections of the paper will summarize the geometric underpinnings and statistical machinery of linear kernel SVMs. All of the major findings and conclusions will be outlined in the next two sections.

## 20 Synopsis of Geometric Underpinnings and Statistical Machinery of Linear Kernel SVMs

This paper has shown that learning linear decision boundaries from training data essentially involves learning the locus of a principal eigenaxis, which has been named a normal eigenaxis. The paper has introduced and developed locus equations of a normal eigenaxis that describe lines, planes, hyperplanes, and normal eigenaxes. The paper has shown that the locus of a normal eigenaxis is an inherent part of any linear curve or surface, where the geometric locus of a linear curve or surface is encoded within the locus of its normal eigenaxis.

This paper has demonstrated how the eigen-coordinate locations of a normal eigenaxis determine the uniform properties exhibited by the points on a linear curve or surface. The paper has shown that normal eigenaxes of linear loci provide exclusive and distinctive reference axes. The paper has also demonstrated that a normal eigenaxis satisfies a linear locus in terms of its eigenenergy, which is the fundamental property of a normal eigenaxis.

This paper has motivated and developed a dual statistical eigenlocus of normal eigenaxis components which encodes the eigen-coordinate locations of an unknown normal eigenaxis of a linear decision boundary. The paper has introduced and developed the constrained primal and the Wolfe dual eigenlocus equations of a strong dual normal eigenlocus, which have been shown to provide a joint statistical estimate of a dual statistical eigenlocus of constrained primal normal eigenaxis components. The paper has demonstrated that finding a separating line, plane, or hyperplane requires estimating the strong dual normal eigenlocus of a linear decision boundary and the bilaterally symmetrical borders which bound it. The paper has also demonstrated that a strong dual normal eigenlocus satisfies a linear decision boundary and the bilaterally symmetrical borders which bound it in terms of a critical minimum eigenenergy.

This paper has demonstrated how the constrained primal and the Wolfe dual eigenlocus equations of a strong dual normal eigenlocus transform two given sets of pattern vectors into a dual statistical eigenlocus of constrained primal normal eigenaxis components, all of which are jointly located in dual and primal, correlated Hilbert spaces, all of which jointly describe correlated linear subspaces of  $\mathbb{R}^N$  and  $\mathbb{R}^d$ , each of which encodes the likelihood of finding an extreme data point in a particular region of  $\mathbb{R}^d$ . Any given dual statistical eigenlocus of constrained primal normal eigenaxis components delineates and satisfies three, symmetrical linear partitioning curves or surfaces in  $\mathbb{R}^d$ .

This paper has demonstrated how all of the constrained primal normal eigenaxis components on a strong dual normal eigenlocus jointly specify a strong dual statistical decision system that delineates bipartite and symmetrical geometric regions of large covariance, located between two data distributions in  $\mathbb{R}^d$ , such that these bipartite, congruent geometric regions delineate (1) bipartite, congruent, non-overlapping geometric regions of large covariance for any two non-overlapping data distributions, or (2) bipartite, congruent geometric regions of data distribution overlap for any two overlapping data distributions. Thereby, this paper has demonstrated the following:

- All of the constrained primal normal eigenaxis components on a strong dual normal eigenlocus jointly specify three, symmetrical hyperplane partitioning surfaces in  $\mathbb{R}^N$  that are interconnected with three, symmetrical linear partitioning curves or surfaces in  $\mathbb{R}^d$ . A strong dual normal eigenlocus delineates and satisfies three, symmetrical linear partitioning curves or surfaces in  $\mathbb{R}^d$ .
- The resultant loci of points on all three linear partitioning curves or surfaces in  $\mathbb{R}^d$  explicitly and exclusively reference the strong dual normal eigenlocus of normal eigenaxis components.
- The geometric loci of a linear decision boundary and its bilaterally symmetrical linear decision borders are regulated by probabilities of finding extreme data points within particular regions of  $\mathbb{R}^d$ .
- Likelihoods encoded within all of the constrained primal normal eigenaxis components on  $\tau_1 - \tau_2$  specify the stochastic behavior of a statistical decision system.
- All of the constrained primal normal eigenaxis components on a strong dual normal eigenlocus jointly encode a robust, data-driven likelihood ratio.
- The fundamental geometric and statistical property exhibited by a constrained primal normal eigenlocus  $\tau = \tau_1 - \tau_2$  is a critical minimum, i.e., a total allowed, eigenenergy.
- The total allowed eigenenergy and the statistical equilibrium point of  $\tau$  are specified by likelihood statistics encoded within correlated normal eigenaxis components on a Wolfe dual normal eigenlocus  $\psi$ .

- The constrained primal normal eigenaxis components on  $\tau$  satisfy a point of statistical equilibrium for which the eigenenergies of the constrained primal normal eigenlocus components  $\tau_1 - \tau_2$  on  $\tau$  are symmetrically balanced with each other in relation to a centrally located statistical fulcrum, which is half of the total allowed eigenenergy of a constrained primal normal eigenlocus  $\tau$ .
- The regularized, data-driven geometric architectures determined by strong dual normal eigenlocus transforms are configured by enforcing joint symmetrical distributions of the eigenenergies of  $\psi$  and  $\tau$  over the eigen-scaled extreme training vectors on  $\tau_1$  and  $\tau_2$ , whereby the eigenenergies of the strong dual normal eigenlocus components  $\tau_1$  and  $\tau_2$  on  $\tau$  are symmetrically balanced with each other.

This paper has shown how the geometric configuration of a Wolfe dual normal eigenlocus in  $\mathbb{R}^N$  effectively determines the geometric configuration of a constrained primal normal eigenlocus in  $\mathbb{R}^d$ . Thereby, this paper has demonstrated the following:

- A Wolfe dual normal eigenlocus delineates and satisfies three, symmetrical hyperplane partitioning surfaces  $H_0$ ,  $H_{+1}$ , and  $H_{-1}$  in  $\mathbb{R}^N$ .
- Geometric configurations of three hyperplane partitioning surfaces  $H_0$ ,  $H_{+1}$ , and  $H_{-1}$  in  $\mathbb{R}^N$  regulate the geometric configurations of three linear partitioning surfaces  $D_0(\mathbf{x})$ ,  $D_{+1}(\mathbf{x})$ , and  $D_{-1}(\mathbf{x})$  in  $\mathbb{R}^d$ .
- Each Wolfe dual normal eigenaxis component on  $\psi$  in  $\mathbb{R}^N$  encodes a maximum covariance estimate in a principal location in  $\mathbb{R}^d$ , which is determined by an eigen-balanced first and second order statistical moment about the geometric locus of an extreme data point in  $\mathbb{R}^d$ .
- Each Wolfe dual normal eigenaxis component on  $\psi$  in  $\mathbb{R}^N$  specifies an eigen-scale for an extreme data point in  $\mathbb{R}^d$ , whereby each constrained primal normal eigenaxis component on  $\tau$  in  $\mathbb{R}^d$  encodes the probability of finding an extreme data point in a particular region of  $\mathbb{R}^d$ .
- The direction of each Wolfe dual normal eigenaxis component on  $\psi$  in  $\mathbb{R}^N$  is identical to the direction of a correlated, constrained primal normal eigenaxis component on  $\tau$  in  $\mathbb{R}^d$ .
- The lengths of each Wolfe dual normal eigenaxis component on  $\psi$  in  $\mathbb{R}^N$  and correlated, constrained primal normal eigenaxis component on  $\tau$  in  $\mathbb{R}^d$  are shaped by identical joint symmetrical distributions of Wolfe dual and constrained primal normal eigenaxis components.
- Each Wolfe dual normal eigenaxis component on  $\psi$  in  $\mathbb{R}^N$  exhibits a length that is shaped by an eigen-balanced pointwise covariance estimate for a correlated extreme training vector in  $\mathbb{R}^d$ , such that the eigenlocus of each constrained primal normal eigenaxis component on  $\tau$  in  $\mathbb{R}^d$  encodes

a maximum covariance estimate in a principal location, in the form of an eigen-balanced first and second order statistical moment about the geometric locus of the extreme point, which describes the probability of finding the extreme data point in a particular region of  $\mathbb{R}^d$ .

- The integrated sum of probabilities encoded within the constrained primal normal eigenaxis components on  $\tau_1$  is balanced with the integrated sum of probabilities encoded within the constrained primal normal eigenaxis components on  $\tau_2$ , so that strong dual normal eigenlocus transforms determine linear decision boundaries and regulated linear decision borders for which class probabilities are equivalent to each other.
- Each constrained primal normal eigenaxis component on  $\tau$  in  $\mathbb{R}^d$  is an eigen-scaled extreme vector that encodes an eigenstate of a statistical decision system which contains a discriminant function that (1) encodes Bayes' likelihood ratio for common covariance training data, and (2) encodes a robust, data-driven likelihood ratio for all other data distributions.

In summary, this paper has demonstrated that strong dual normal eigenlocus transforms generate robust statistical decision systems for a wide variety of data distributions, including completely overlapping distributions. This paper has also demonstrated how properly regularized linear kernel SVMs implement strong dual normal eigenlocus transforms.

## 21 Conclusions

The problem of learning linear decision boundaries for overlapping sets of data has been resolved. This long-standing problem was generally deemed insoluble. The dilemma has been resolved by means of a dual statistical eigenlocus of normal eigenaxis components, i.e., a strong dual normal eigenlocus of eigen-scaled extreme data points, all of which encode a robust likelihood ratio, where all of the eigen-scaled extreme data points sit on and satisfy the strong dual normal eigenlocus, and all of the points on a statistical decision system of symmetrical linear partitioning curves or surfaces explicitly and exclusively reference the strong dual normal eigenlocus.

The discoveries presented in this paper specify effective designs for linear kernel SVMs. The discoveries also define a statistical model for linear kernel SVM that represents the relevant aspects of probabilistic, binary, linear classification systems.

To summarize, an estimation process has been introduced that provides consistent fits of random data points to unknown normal eigenaxis components of linear decision boundaries. A computer-implemented method has been formulated that transforms two sets of pattern vectors, generated by any two probability distributions whose expected values and covariance structures do not vary significantly over time, into a dual statistical eigenlocus of normal eigenaxis components, all of which are jointly and symmetrically located in correlated,

dual and primal Hilbert spaces, all of which jointly describe correlated, linear subspaces of  $\mathbb{R}^N$  and  $\mathbb{R}^d$ , all of which encode a robust likelihood ratio, all of which jointly specify a statistical decision system that delineates a bipartite, symmetric partitioning of a region of large covariance between two overlapping or non-overlapping data distributions in  $\mathbb{R}^d$ .

The statistical decision system provides a building block for probabilistic, multiclass linear classifiers, a probabilistic, binary linear classifier for overlapping and non-overlapping data distributions, an optimal, probabilistic, binary linear classifier for common covariance data, a statistical gauge for data distribution overlap and Bayes' error, and a statistical gauge that is a definitive indicator of homogeneous data distributions.

An upcoming paper will introduce an estimation process that provides consistent fits of random data points to unknown principal eigenaxis components of unknown second-order decision boundaries that take the form of  $d$ -dimensional circles, ellipses, hyperbolae, and parabolas. The discoveries presented in the paper are expected to specify effective designs for polynomial kernel SVMs. The discoveries are also expected to define a statistical model for polynomial kernel SVM that represents the relevant aspects of probabilistic, binary, nonlinear classification systems.

## Acknowledgment

The author is indebted to Oscar Gonzalez and Garry Jacyna. The author's master's thesis Reeves [1995] was the primary impetus for this work. The counsel of Oscar Gonzalez motivated and sustained the trailblazer within the author. Some of the material in this paper includes portions of the author's PhD dissertation Reeves [2009]. Part of this work would not have occurred without the support of the author's adviser at the MITRE Corporation. The guidance of Garry Jacyna enabled the author to successfully navigate the PhD pipeline at George Mason University.

## References

- C. Ash. *The Probability Tutoring Book*. IEEE Press, 1993.
- I. Asimov. *Understanding Physics: Motion, Sound, and Heat*. Mentor, 1966.
- A. Barron, J. Rissanen, and B. Yu. The minimum description length principle in modeling and coding. *Information Theory, IEEE Transactions on*, 44(6): 4–37, Oct. 1998.
- K. P. Bennett and C. Campbell. Support vector machines: Hype or hallelujah? *SIGKDD Explorations*, 2(2):1–13, 2000.
- B. Boser, I. Guyon, and V. Vapnik. A training algorithm for optimal margin classifiers. In *Proceedings of the 5th Annual ACM Workshop on Computational Learning Theory*, pages 144–152. ACM Press, 1992.

- L. Breiman. Statistical modeling: The two cultures. *Statistical Science*, 16(3): 199–231, 1991.
- C. Burges. A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery*, 2:121–167, 1998.
- H. Byun and S. Lee. Applications of support vector machines for pattern recognition: A survey. *LNCS*, 2388:213–236, 2002.
- V. Cherkassky and F. Mulier. *Learning From Data: Concepts, Theory and Methods*. Wiley-Interscience, 1998.
- P. Cooper. The hyperplane in pattern recognition. *Cybernetica*, 4:215–238, 1962.
- C. Cortes and V. Vapnik. Support-vector networks. *Machine Learning*, 20(3): 273–297, 1995.
- T. Cover. Geometrical and statistical properties of systems of linear inequalities with applications in pattern recognition. *IEEE Transactions on Electronic Computers*, pages 326–334, 1965.
- N. Cristianini and J. Shawe-Taylor. *An Introduction to Support Vector Machines and Other Kernel-based Learning Methods*. Cambridge University Press, 2000.
- C. Daniel and F. Wood. *Fitting Equations to Data*. John Wiley and Sons, 1979.
- P. J. Davis. *Interpolation and Approximation*. Blaisdell, 1963.
- P. J. Davis. *The Mathematics of Matrices*. John Wiley and Sons, 1973.
- R. Duda, P. Hart, and D. Stork. *Pattern Classification*. John Wiley and Sons, 2001.
- L. Eisenhart. *Coordinate Geometry*. Dover Publications, 1939.
- T. Eitrich and B. Lang. Efficient optimization of support vector machine learning parameters for unbalanced datasets. *J COMPUT APPL MATH*, 196(2): 425–436, 2006.
- W. H. Engl, M. Hanke, and A. Neubauer. *Regularization of Inverse Problems*. Kluwer Academic Publishers, 2000.
- R. Fletcher. *Practical Methods of Optimization*. Wiley, 2000.
- B. Flury. *A First Course in Multivariate Statistics*. Springer-Verlag, 1997.
- K. Fukunaga. *Introduction to Statistical Pattern Recognition*. Academic Press, 1990.
- S. Geman, E. Bienenstock, and R. Doursat. Neural networks and the bias/variance dilemma. *Neural Computation*, 4:1–58, 1992.

- N. Gershenfeld. *The Nature of Mathematical Modeling*. Cambridge University Press, 1999.
- M. A. Goldberg. A method of adjoints for ill-posed equations of the first kind. *Applied Mathematics and Computation*, 5:123–130, 1979.
- C. Groetsch. *The Theory of Tikhonov Regularization for Fredholm Equations of the First Kind*. Pitman Advanced Publishing Group, 1984.
- C. Groetsch. *Inverse Problems in the Mathematical Sciences*. Vieweg, 1993.
- P. C. Hansen. *Rank-Deficient and Discrete Ill-Posed Problems*. SIAM, 1998.
- T. Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning*. Springer, 2001.
- S. Haykin. *Neural Networks and Learning Machines*. Prentice Hall, 2009.
- S. Hewson. *A Mathematical Bridge*. World Scientific Publishing, 2009.
- J. Hillman. *Kinds of Power*. Random House, 2012.
- A. E. Hoerl. Application of ridge analysis to regression problems. *Chem Eng Prog*, 58:58–60, 1962.
- S. M. Kay. *Fundamentals of Statistical Signal Processing: Estimation Theory*. Prentice-Hall, 1993.
- J. Keener. *Principles of Applied Mathematics, Transformation and Approximation*. Perseus Books, 2000.
- P. Lancaster and K. Salkauskas. *Curve and Surface Fitting*. Academic Press, 1986.
- B. P. Lathi. *Signal Processing and Linear Systems*. Berkley-Cambridge, 1998.
- D. C. Lay. *Linear Algebra and Its Applications*. Addison Wesley, 2006.
- Y. Liang, Q. Xu, H. Li, and D. Cao. *Support Vector Machines and Their Application in Chemistry and Biotechnology*. CRC Press, 2011.
- P. Linz. *Theoretical Numerical Analysis*. Dover Publications, 1979.
- P. Linz and R. Wang. *Exploring Numerical Methods*. Jones and Bartlett Publishers, 2003.
- D. Luenberger. *Optimization by Vector Space Methods*. John Wiley and Sons, 1969.
- D. Luenberger. *Linear and Nonlinear Programming*. Kluwer Academic Publishers, 2003.



- D. Meadows. *Thinking in Systems: A Primer*. Chelsea Green Publishing Company, 2008.
- A. Mertins. *Signal Analysis*. John Wiley and Sons, 1999.
- C. Meyer. *Matrix Analysis and Applied Linear Algebra*. SIAM, 2000.
- T. Mitchell. *Machine Learning*. McGraw-Hill Company, 1997.
- T. K. Moon and W. C. Stirling. *Mathematical Methods and Algorithms for Signal Processing*. Prentice-Hall, 2000.
- S. Nash and A. Sofer. *Linear and Nonlinear Programming*. McGraw-Hill Company, 1996.
- A. Naylor and G. Sell. *Linear Operator Theory in Engineering and Science*. Holt Rinehart and Winston, 1971.
- E. Nichols. *Analytic Geometry*. D. C. Heath and Company, 1893.
- M. Rahman. *Applied Numerical Analysis*. WIT Press, 2004.
- S. Rao. *Applied Numerical Methods for Engineers and Scientists*. Prentice Hall, 2002.
- D. M. Reeves. Generalization metrics for neural modeling applications in system identification. Master's thesis, Old Dominion University, 1995.
- D. M. Reeves. Finding the fittest separating hyperplane for the case of common covariance matrices. In *Joint Statistical Meetings*, 2007.
- D. M. Reeves. *Properly Specified Functional Mappings and Support Vector Learning Machines*. PhD thesis, George Mason University, 2009.
- D. M. Reeves and G. M. Jacyna. Support vector machine regularization. *WIREs Computational Statistics*, 3:204–215, 2011.
- B. Scholkopf and A. Smola. *Learning with Kernels*. MIT Press, 2002.
- J. Stewart. *Multivariable Calculus: Concepts and Contexts*. Cengage Learning, 2009.
- G. Strang. *Introduction to Applied Mathematics*. Wellesley-Cambridge Press, 1986.
- J. L. Synge. *The Hypercircle in Mathematical Physics*. Cambridge University Press, 1957.
- Tanner and Allen. *Analytic Geometry*. American Book Company, 1898.
- A. Tikhovov and V. Arsenin. *Solutions of Ill-Posed Problems*. John Wiley and Sons, 1977.

- L. N. Trefethen and D. Bau. *Numerical Linear Algebra*. SIAM, 1998.
- H. VanTrees. *Detection, Estimation, and Modulation Theory: Part I*. John Wiley and Sons, 1968.
- G. Wahba. *Inverse and Ill-Posed Problems*. Academic Press, 1987.
- A. Whitehead. *An Introduction to Mathematics*. Henry Holt and Company, 1911.
- M. Zhdanov. *Geophysical Inverse Theory and Regularization Problems*. Elsevier Science, 2002.